



Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions

Guillaume Gales

► To cite this version:

Guillaume Gales. Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions. Interface homme-machine [cs.HC]. Université de Toulouse, 2011. Français. <tel-00622859>

HAL Id: tel-00622859

<https://tel.archives-ouvertes.fr/tel-00622859>

Submitted on 12 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse III - Paul Sabatier
Discipline ou spécialité : Informatique

Présentée et soutenue par Guillaume GALES
Le 8 juillet 2011

Titre :

*Mise en correspondance de pixels pour la stéréovision binoculaire par propagation
d'appariements de points d'intérêt et sondage de régions*

JURY

Rapporteurs : M. Adrien BARTOLI
M. David FOFI

Examineur : M. Frédéric JURIE

Directeur de thèse : M. Patrice DALLE
Encadrants : M^{me} Sylvie CHAMBON
M. Alain CROUZIL

Professeur Université d'Auvergne, Clermont-Ferrand
Professeur Université de Bourgogne, Le Creusot
Professeur Université de Caen Basse-Normandie,
Caen
Professeur Université Paul Sabatier, Toulouse
Chargée de recherche IFSTTAR, Nantes
Maître de conférences Université Paul Sabatier,
Toulouse

École doctorale : Mathématiques, Informatique et Télécommunications de Toulouse
Unité de recherche : IRIT- UMR 5505
Directeur de Thèse : M. Patrice DALLE

Résumé

LA mise en correspondance stéréoscopique est un des principaux problèmes de la vision par ordinateur. Elle consiste à trouver dans deux images d'une même scène, prises à des positions différentes, les couples de pixels qui sont les projections d'un même point de la scène. Durant ces vingt dernières années, de nombreuses méthodes *locales* et *globales* ont été proposées pour résoudre ce problème. Plus récemment, grâce à un protocole d'évaluation de référence dans la communauté, une catégorie de méthodes *fondées sur les régions* a montré des performances intéressantes dans le cadre de la stéréovision binoculaire dite « *small-baseline* » (dont les images sont prises à des positions proches). Une contribution de ce travail porte sur un type particulier de méthode locale appelé *propagation de germes*. La zone de recherche des correspondants est réduite aux voisinages d'appariements fiables appelés *germes*, en faisant l'hypothèse que, généralement, deux pixels voisins ont des correspondants proches. Cela permet de réduire les temps de calculs et d'éviter certaines ambiguïtés. Cependant, le succès de ce type de méthode est fortement dépendant du choix de ces germes. Dans ce mémoire, nous proposons une étude de l'étape de sélection des germes. Nous nous concentrons sur la mise en correspondance de points d'intérêt. Il s'agit de points de l'image ayant des caractéristiques intéressantes pour une application donnée. Dans notre cas, nous avons besoin de pixels qui peuvent être mis en correspondance de manière sûre. Nous comparons quatorze détecteurs bien connus associés à cinq mesures de corrélation. Certaines de ces mesures sont conçues pour être robustes à un des principaux problèmes de la mise en correspondance stéréoscopique : les ruptures de profondeur. En outre, cette étude donne des conseils sur la manière de choisir les paramètres des différentes méthodes afin de trouver les meilleurs germes possibles selon des critères donnés. Ensuite, ces germes sont utilisés avec deux approches de propagation et les résultats sont évalués.

Une autre contribution porte sur une nouvelle approche de mise en correspondance stéréoscopique dense fondée sur les régions. Différentes segmentations couleur sont utilisées. Ensuite, plusieurs instances d'un modèle de surface sont calculées pour les différentes régions selon des disparités initiales tirées au sort. Pour chaque pixel, chaque instance donne une disparité qui est considérée comme un vote. Finalement, la disparité qui reçoit le plus de voix est sélectionnée comme disparité finale. Cette approche est relativement simple à implémenter et s'avère efficace en donnant des résultats compétitifs vis-à-vis de l'état de l'art.

Mots-clés : mise en correspondance stéréoscopique, propagation de germes, détection de points d'intérêt, sondage de régions.

Abstract

STEREO matching is one of the main topics in computer vision. It consists in finding in two images of a same scene, taken from different viewpoints, the pairs of pixels which are the projections of a same scene point. Since the last twenty years, many *local* and *global* methods have been proposed to solve this problem. More recently, according to a reference evaluation protocol in the community, *region-based* methods showed interesting result in small-baseline binocular stereo (where images are taken nearby).

The idea is to apply a colour segmentation algorithm on the images assuming that each pixel within a segment belongs to a same object surface. Then, the parameters of a surface model are computed, in the disparity space, for each segment according to initial disparities usually computed with a local method. Finally, a global optimization is performed to refine the results.

A contribution of this thesis deals with a special kind of local method called *seeds propagation*. The search area of a correspondent is reduced to the neighbourhoods of reliable matches called *seeds*. This can help to reduce the computation time and to avoid some ambiguities. However, the succes of such a method depends on the choice of these seeds. In this dissertation, we give a study of the seeds selection step. We focus on feature points matching. These are special points in the image with interesting characteristics for a given application. In our case, we need pixels that can be matched with high confidence. We compare fourteen different well-known detectors linked to five correlation measures. Some of these measures are meant to be robust to one of the main challenge in stereo matching: depth discontinuities. Besides, this study gives advice on the choice of the parameters of the different techniques to be able to find the best solutions according some given criteria. Then, these seeds are used with two approaches of propagation and the results are evaluated.

Another contribution deals with a new region-based approach for dense stereo matching. Different colour segmentations are used. Then, many instances of a surface model are computed for the different regions according to initial disparities selected randomly. For each pixel, each instance gives a disparity value regarded as a vote. Finally, the most voted value is selected as the final disparity. This approach is relatively easy to implement and very effective giving competitive results among the state of the art.

Keywords: stereo matching, seeds propagation, feature point detection, randomize voting scheme.

Remerciements

Je remercie Adrien Bartoli et David Fofi d'avoir accepté d'être rapporteurs de ce travail, ainsi que Frédéric Jurie d'avoir accepté d'être examinateur et président du jury de cette thèse. J'ai particulièrement apprécié les remarques constructives dont vous m'avez fait part. Je remercie également Patrice Dalle d'avoir accepté d'être directeur de thèse.

Je remercie particulièrement Alain Crouzil et Sylvie Chambon pour leur investissement, la pertinence de leurs remarques et leur patience dans l'encadrement de mes travaux.

Je remercie également : Adrian, Agathe, Andra, Anne-Marie, Anne-Marie, Anthony, Aurélie, Axel, Benoît, Blandine, Carole, Céline, Christophe, Cyril, Denis, Énora, Fabien, François, Françoise, Françoise, Frédéric, Frédéric, Guillaume, Isabelle, Ivan, Jean-Baptiste, Jean-Claude, Jean-Denis, Jean-Marc, Jean-Philippe, Jean-Pierre, Jean-Pierre, Jérôme, Jérôme, Jonathan, Julie, Juliette, Laure, Laurie, Lilian, Loïc, Lucille, Marie, Marion, Martine, Mathieu, Mathieu, Matilde, Mélanie, Mikaël, Monia, Naïma, Olivier, Olivier, Philippe, Philippe, Pierre, Rémi, Rémi, Renaud, Rodolphe, Sabine, Samir, Sylvain, Théo, Timéo, Vincent et Viorica.

Table des matières

Résumé	iii
Abstract	v
Table des matières	xi
Table des figures	xv
Liste des tableaux	xviii
Liste des algorithmes	xix
1 Introduction	21
1.1 Mise en correspondance de pixels	22
1.2 Principales difficultés de la mise en correspondance de pixels	22
1.3 Problématique	26
1.4 Organisation du mémoire	26
2 Mise en correspondance de pixels pour la stéréovision binoculaire	29
2.1 Notations	30
2.2 Méthodes locales	33
2.2.1 Mise en correspondance locale classique : algorithme élémentaire . .	33
2.2.2 Ajout de contraintes	35
2.2.3 Prise en compte des difficultés de la mise en correspondance	41
2.2.4 Mise en correspondance par propagation de germes	51
2.3 Méthodes globales	57
2.3.1 Modélisation du problème à l'aide des champs de Markov	59
2.3.2 Fonctions de coût	61
2.3.3 Méthodes d'optimisation	63

TABLE DES MATIÈRES

2.4	Méthodes fondées sur les régions	68
2.5	Conclusion	72
3	Sélection des germes initiaux pour la mise en correspondance par propagation	75
3.1	Détection de points d'intérêt	76
3.1.1	Famille dérivée	81
3.1.2	Famille morphologique	91
3.1.3	Famille entropique	95
3.1.4	Prise en compte de la couleur	96
3.1.5	Synthèse des détecteurs	97
3.2	Appariement de points d'intérêt pour la sélection de germes	97
3.3	Conclusion	100
4	Évaluation des méthodes de sélection de germes initiaux par appariement de points d'intérêt	101
4.1	Critères d'évaluation	104
4.2	Détermination expérimentale des paramètres de sélection des germes	107
4.2.1	Protocole expérimental	107
4.2.2	Résultats	113
4.3	Complémentarité des germes	123
4.4	Conclusion	128
5	Mise en correspondance par propagation multi-mesure de germes	133
5.1	Algorithme de propagation multi-mesure de germes	134
5.1.1	Sélection des germes	134
5.1.2	Propagation multi-mesure	135
5.2	Critères d'évaluation	136
5.3	Évaluation de la propagation multi-mesure	138
5.3.1	Protocole expérimental	138
5.3.2	Comparaison avec la mise en correspondance locale classique	146
5.4	Conclusion	150
6	Mise en correspondance par sondage de régions	153
6.1	Algorithme proposé	154
6.2	Évaluation du sondage de régions	158
6.2.1	Protocole expérimental	158
6.2.2	Résultats	160
6.3	Conclusion	161

7	Conclusions et perspectives	165
7.1	Bilan des méthodes de mise en correspondance	165
7.2	Bilan des contributions	166
7.3	Perspectives	167
	Bibliographie	167

Table des figures

1.1	Couple d'images et anaglyphe.	22
1.2	Modèle géométrique du capteur stéréoscopique binoculaire.	23
1.3	Zones occultées.	24
1.4	Zones de discontinuité de profondeur.	24
1.5	Zones homogènes.	25
1.6	Zones à textures répétitives.	25
2.1	Familles de méthodes de mise en correspondance de pixels.	30
2.2	Vecteur de disparité.	32
2.3	Couple d'images et carte de disparité.	32
2.4	Mise en correspondance locale.	34
2.5	Géométrie épipolaire.	36
2.6	Rectification épipolaire.	36
2.7	Contrainte d'ordre.	38
2.8	Contrainte d'unicité.	38
2.9	Contrainte de symétrie et de symétrie faible.	39
2.10	Contrainte de rang.	41
2.11	Mesure d'ambiguïté.	42
2.12	Mesure d'imprécision.	42
2.13	Multi-fenêtrage fixe.	47
2.14	Multi-fenêtrage par croissance.	47
2.15	Multi-fenêtrage adjacent.	48
2.16	Fenêtre adaptative.	49
2.17	Principe de la propagation.	52
2.18	Approche simultanée de la propagation.	54
2.19	Approche séquentielle de la propagation.	54
2.20	Contrainte de région.	57
2.21	Principe des méthodes globales.	58

TABLE DES FIGURES

2.22	Évolution de la carte de disparité du couples <i>Cones</i> durant le processus de la mise en correspondance globale.	58
2.23	Modélisation du problème à l'aide des champs de Markov.	61
2.24	Modèles de fonction de lissage.	64
2.25	Étapes des méthodes de mise en correspondance fondées sur les régions. . .	70
2.26	Exemple de segmentation par <i>mean-shift</i>	73
2.27	Espace de classification 3D.	74
3.1	Répétabilité	77
3.2	Exemples d'échelles différentes.	80
3.3	Détecteur de Moravec.	82
3.4	Détecteur MIC.	85
3.5	SIFT – Différences de gaussiennes et sélection des extrema.	88
3.6	Relocalisation.	89
3.7	Image intégrale.	91
3.8	<i>Box filtering</i>	92
3.9	Détecteur SUSAN.	94
3.10	Détecteur FAST.	95
4.1	Précision de l'erreur d'appariement.	104
4.2	Taux de rappel.	105
4.3	Segmentations utilisées pour calculer le taux de répartition.	106
4.4	Taux de répartition.	107
4.5	Recuit simulé.	114
4.6	Classement des couples détecteur-corrélation par compromis A.	117
4.7	Vue d'ensemble des résultats pour le compromis A.	123
4.8	Coefficient de cardinalité des germes corrects.	127
4.9	Pixels gauches des germes des 6 meilleurs ensembles.	131
5.1	Exemple de disparités obtenues par une mesure classique puis robuste. . .	134
5.2	Complètement et validation des germes.	136
5.3	Carte des zones difficiles.	137
5.4	Segmentations utilisées pour la contrainte de région lors des propagations. .	141
5.5	Taux d'acceptation en fonction de la densité des propagations (coloration par approche).	146
5.6	Taux d'acceptation en fonction de la densité des propagations (coloration par contrainte de région).	146
5.7	Taux d'acceptation en fonction de la densité des propagations (coloration par mesure).	147
5.8	Cartes des disparités propagées.	148
5.9	Cartes de disparité obtenues par mise en correspondance locale classique. .	149
6.1	Étapes de la mise en correspondance par sondage de régions.	156
6.2	Principe du sondage de régions.	157

6.3	Carte et histogrammes des valeurs de confiance.	159
6.4	Cartes de disparité obtenues par sondage des régions.	162

Liste des tableaux

3.1	Détecteurs de la famille dérivée.	98
3.2	Détecteurs de la famille morphologique.	99
3.3	Détecteur de la famille entropique.	99
4.1	Images testées.	109
4.2	Détecteurs de point d'intérêt comparés.	110
4.3	Mesures de corrélation comparées.	110
4.4	Classement des couples détecteur-corrélation par rang sur le compromis A.	117
4.5	Paramètres optimaux pour le compromis A avec FAST-CENSUS.	118
4.6	Paramètres médians optimaux pour le compromis A.	118
4.7	Échelles médianes des détecteurs à échelle fixe.	119
4.8	Seuils médians des détecteurs.	120
4.9	Tailles médianes des fenêtres de corrélation.	120
4.10	Seuils médians des mesures de corrélation.	121
4.11	Reclassement des couples détecteur-corrélation par rang de compromis A.	122
4.12	Classement par rang médian des meilleures paires de couples détecteur-corrélation (piste 1).	125
4.13	Reclassement des couples détecteur-corrélation par compromis A (tri par taux de rappel).	126
4.14	Classement des couples détecteur-corrélation par rang de compromis B.	128
4.15	Paramètres des meilleurs couples détecteur-corrélation pour obtenir le meilleur compromis B.	129
4.16	Classement des meilleures paires de couples détecteur-corrélation (piste 3).	129
4.17	Reclassement des couples détecteur-corrélation par compromis B (tri par taux de rappel).	130
4.18	Performances des ensembles de germes complémentaires.	132
5.1	Ensembles de germes testés.	139
5.2	Taux des ensembles de germes évalués.	142
5.3	Classement par compromis C des meilleures propagations.	145

LISTE DES TABLEAUX

5.4	Résultats de la mise en correspondance locale classique.	149
5.5	Bilan des méthodes locales.	151
6.1	Résultats de la mise en correspondance par sondage de régions.	160
6.2	Pourcentages moyens de mauvais appariements.	161
7.1	Bilan des méthodes de mise en correspondance.	165
7.2	Bilan des résultats issus de l'étude proposée.	167

Liste des algorithmes

2.1	Algorithme élémentaire de mise en correspondance locale classique.	34
2.2	Approche simultanée de la propagation.	55
2.3	Approche séquentielle de la propagation.	56
2.4	Algorithme de Ford et Fulkerson.	65
2.5	Algorithme de Veksler.	66
2.6	Recuit simulé.	67
2.7	Algorithme génétique.	67
4.1	Recuit simulé pour l'estimation des paramètres de la sélection des germes.	115

1 Introduction

L'INFORMATIQUE est définie comme étant la science du traitement automatique de l'information. La vision par ordinateur qui consiste à traiter l'information contenue dans des images numériques s'inscrit donc dans ce cadre. Ces images numériques sont constituées de *pixels* (*picture element*) qui sont généralement arrangés sous la forme d'une grille représentée par une matrice. Chaque pixel a donc une position dans l'image et est associé à une information sur le rayonnement lumineux reçue par le capteur (intensité ou couleur par exemple).

En stéréovision binoculaire, nous travaillons avec un couple d'images d'une même scène prises à des positions différentes de façon analogue au système de vision humain. Les caméras sont semblables aux yeux qui captent les rayons lumineux provenant de la scène et, comme en vision humaine, c'est le décalage entre les deux points de vue qui permet d'estimer le relief. Un exemple illustrant ce principe est l'anaglyphe de la figure 1.1 où les vues gauche et droite sont superposées. La vue gauche est teintée en rouge et la vue droite en cyan. Ainsi, si le lecteur dispose d'une paire de lunettes où un filtre rouge est placé devant la lunette gauche, et un filtre cyan devant la lunette droite, l'œil gauche voit l'image gauche et l'œil droit voit l'image droite. Le cerveau donne au lecteur une impression de profondeur. Le rôle de la stéréovision binoculaire est quant à lui comparable au traitement réalisé par le cerveau pour retrouver la profondeur. Il s'agit alors de calculer dans la scène la position des éléments qui se projettent en chaque pixel.

Quatre étapes sont alors nécessaires pour retrouver l'information de profondeur :

- **Acquisition** – Les deux images de la même scène sont acquises au même moment par deux caméras différentes ou bien, dans le cas d'une scène fixe, une seule caméra peut être utilisée puis déplacée pour obtenir la seconde image.
- **Calibrage** – Cette étape consiste à déterminer les paramètres internes et externes du modèle géométrique du capteur stéréoscopique qui est décrit sur la figure 1.2. Le lecteur intéressé par cette étape trouvera notamment plus d'information dans [Bocquillon 08].
- **Mise en correspondance de pixels** – Il s'agit de retrouver sur les deux images les paires de pixels qui correspondent à la projection d'un même élément de la scène, voir

la figure 1.2.

- **Reconstruction 3D** – Elle consiste à calculer pour chaque pixel la position dans l'espace du point qui s'est projeté en ce pixel.

Les travaux présentés dans ce mémoire concernent l'étape de mise en correspondance de pixels pour la stéréovision binoculaire où les deux points de vue sont relativement proches (*small-baseline*). Cette thèse s'inscrit dans la continuité des travaux présentés dans [Chambon 05].

1.1 Mise en correspondance de pixels

L'objectif est de retrouver sur les deux images les paires de pixels qui correspondent à la projection d'une même entité. Il s'agit d'un problème largement étudié qui présente différentes difficultés que nous détaillons ci-dessous. De nombreuses techniques ont été proposées pour tenter de résoudre certains de ces problèmes, sachant qu'il est difficile de tous les prendre en compte en même temps, mais qu'il est possible de trouver des compromis.



FIGURE 1.1: *Couple d'images et anaglyphe.*

1.2 Principales difficultés de la mise en correspondance de pixels

Les méthodes de mise en correspondance de pixels s'appuient généralement sur l'hypothèse que les voisinages de deux correspondants se ressemblent. Les principales difficultés rencontrées lors d'une mise en correspondance sont dues essentiellement à deux raisons :

- **Information manquante** – L'information qui permettrait de réaliser une mise en correspondance de façon certaine n'est pas présente dans l'image. Il est donc difficile de réaliser la mise en correspondance sans caractéristique discriminante complémentaire. C'est le cas notamment dans les zones occultées, les zones homogènes, les zones de discontinuité de profondeur et de raccourcissement.
 - ▶ *Zones occultées* – Il s'agit des zones dans lesquelles les éléments de la scène visibles sur une image n'apparaissent pas sur l'autre image car, vus depuis un autre angle, ces éléments sont cachés par d'autres, voir figure 1.3.
 - ▶ *Zones de discontinuité de profondeur* – Il s'agit des zones situées aux frontières des objets ayant des profondeurs différentes. Cela peut se traduire par des voisinages de deux correspondants qui ne se ressemblent pas nécessairement, voir figure 1.4.
 - ▶ *Zones de raccourcissement* – Différents points de la scène se projettent en plusieurs pixels d'une image mais sur un même pixel de l'autre image. Cela peut se produire lorsqu'une surface de la scène est très inclinée par rapport au plan image.

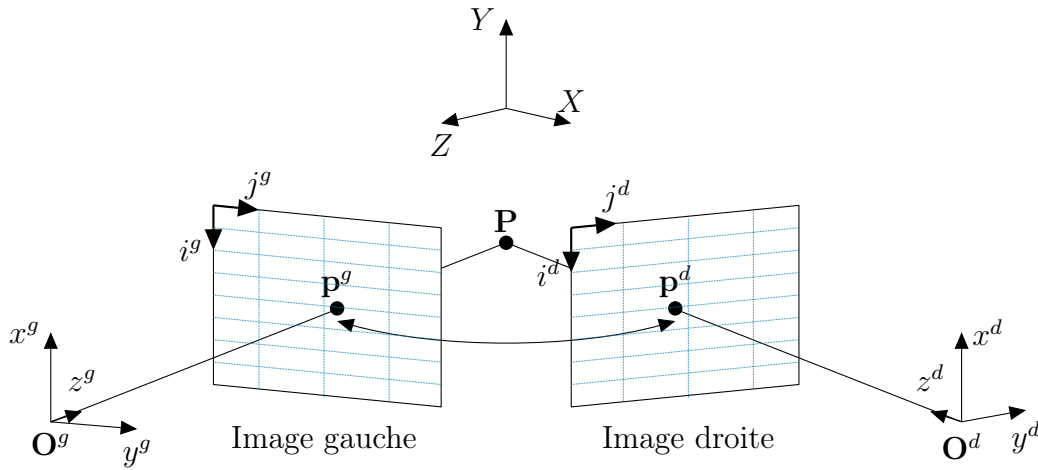


FIGURE 1.2: *Modèle géométrique du capteur stéréoscopique binoculaire.* – Les pixels notés \mathbf{p}^g et \mathbf{p}^d se correspondent car ils sont les projections du même point de la scène \mathbf{P} sur les images gauche et droite. Les repères (x^g, y^g, z^g) et (x^d, y^d, z^d) sont respectivement les repères caméras gauche et droite. Le repère (X, Y, Z) est le repère scène. Les repères images sont notés (i^g, j^g) et (i^d, j^d) .

- **Information ambiguë** – Nous pouvons considérer que l'information est ambiguë lorsqu'elle est présente dans l'image mais qu'elle ne permet pas de distinguer les correspondants.
 - ▶ *Zones homogènes* – Tous les pixels d'une zone homogène se ressemblent, au bruit près. Par conséquent il est difficile de les distinguer, voir figure 1.5.
 - ▶ *Zones à textures répétitives* – Il existe une ambiguïté lorsqu'un même motif se répète plusieurs fois, voir figure 1.6.

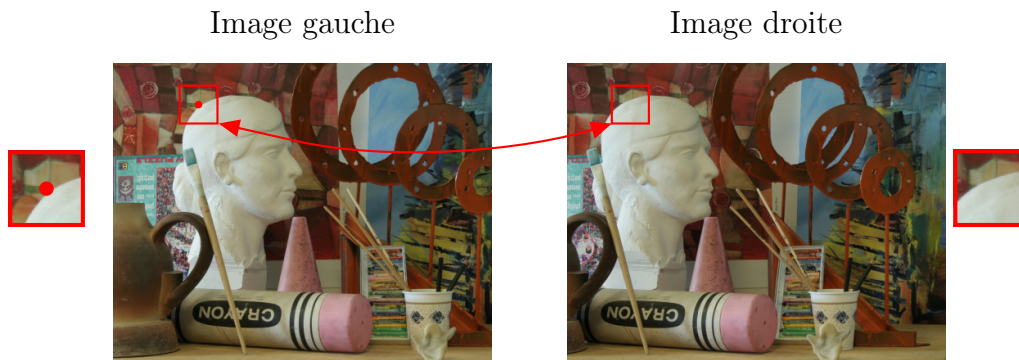


FIGURE 1.3: *Zones occultées.* – Le pixel marqué d'un point rouge a son correspondant occulté dans l'image droite.

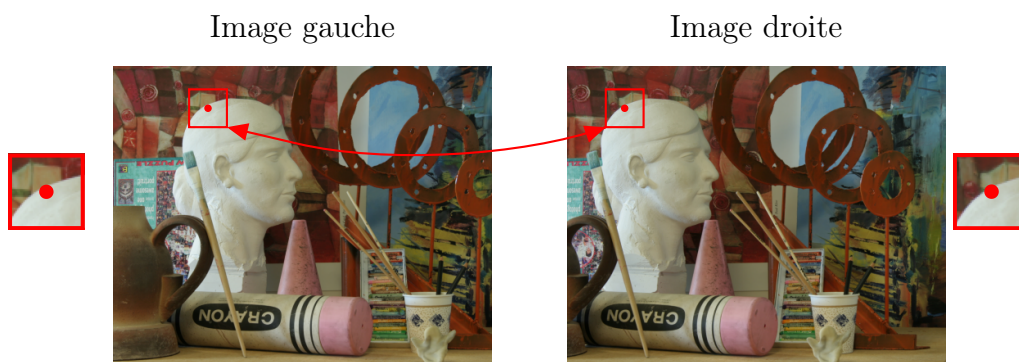


FIGURE 1.4: *Zones de discontinuité de profondeur.* – Les deux pixels marqués d'un point rouge se correspondent mais leurs voisinages diffèrent car ils sont proches d'une discontinuité de profondeur.

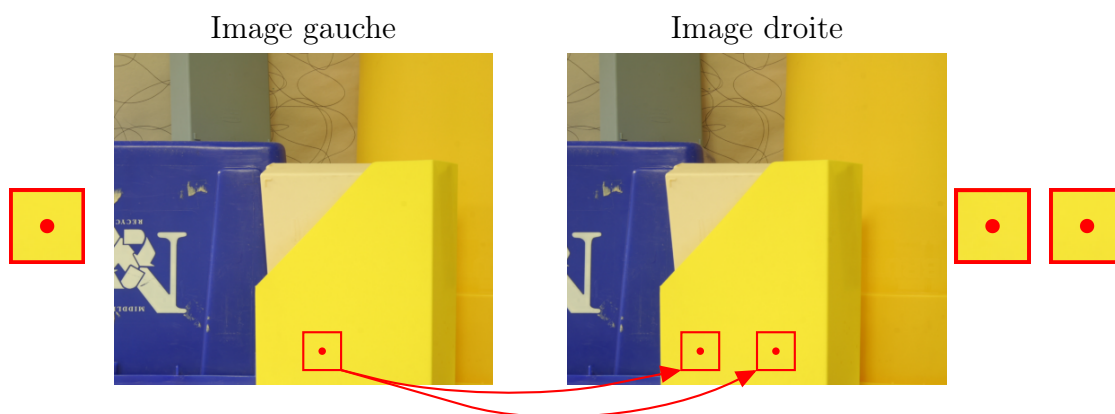


FIGURE 1.5: Zones homogènes. – Dans les zones homogènes, il y a peu d’information permettant de caractériser les pixels ; il est alors difficile de distinguer le correspondant.

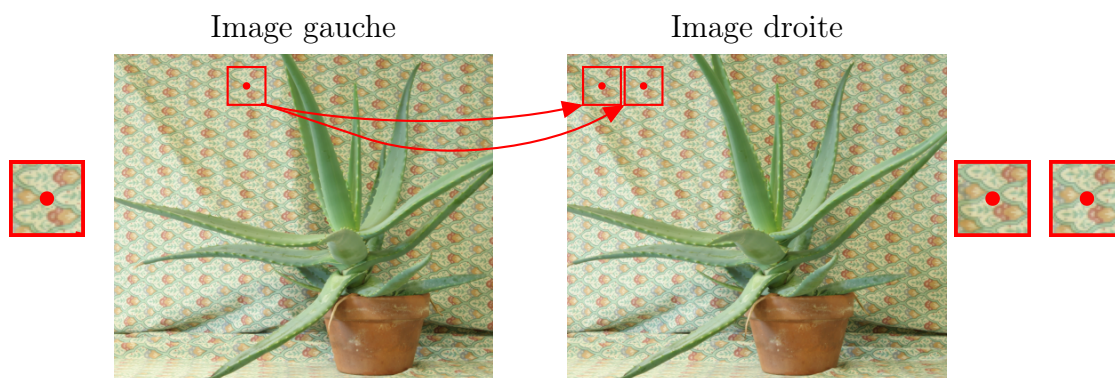


FIGURE 1.6: Zones à textures répétitives. – Pour un pixel, il existe plusieurs candidats possibles dont les voisinages se ressemblent. Ainsi, il est difficile de distinguer le bon correspondant.

1.3 Problématique

Les méthodes de mise en correspondance de pixels peuvent être regroupées en deux familles. La première famille est celles des méthodes *globales* qui prennent en compte la totalité de l'image et cherchent à minimiser une fonction de coût qui exprime les erreurs de mise en correspondance. La seconde famille est celle des méthodes *locales* où les paires de pixels correspondants sont trouvées en mesurant la similarité de leurs voisinages. Le correspondant de chaque pixel est recherché parmi un ensemble de candidats situés dans une zone de recherche bien délimitée.

Les méthodes locales sont relativement faciles à mettre en œuvre et sont encore largement utilisées dans différentes applications. Malheureusement, elles sont sensibles aux difficultés présentées au paragraphe précédent. Cependant, il existe une sous-famille particulière de méthodes locales par *propagation de germes*. Elles permettent *a priori* de réduire certaines ambiguïtés auxquelles sont confrontées les méthodes classiques ainsi que les temps de calcul. Pourtant, ces méthodes ont été peu exploitées. L'un des objectifs du travail présenté dans ce mémoire est de mener une étude détaillée sur cette piste de recherche.

Par ailleurs, le classement donné par le protocole d'évaluation des méthodes de mise en correspondance dense de pixels pour la stéréovision binoculaire de Middlebury¹ [Scharstein 02] met en évidence les très bons résultats obtenus par les méthodes *fondées sur les régions*. Il s'agit d'une famille de méthodes hybrides car elles font généralement intervenir à la fois une étape de mise en correspondance locale pour initialiser l'algorithme et une étape d'optimisation globale pour raffiner les résultats. Les méthodes fondées sur les régions s'appuient sur l'hypothèse que les pixels d'une même région issue d'une segmentation en couleur homogène sont les projections d'une même surface. Cette hypothèse est généralement vérifiée presque partout et les contours des objets de la scène, où se situent des discontinuités de profondeur, coïncident avec les frontières des régions. Un autre objectif de ce travail consiste à étudier la manière de prendre en compte ce constat et d'évaluer l'apport de l'utilisation des régions dans la qualité des résultats de la mise en correspondance.

1.4 Organisation du mémoire

Ce mémoire s'articule autour des chapitres suivants :

- **Chapitre 2** – Il s'agit de l'état de l'art des méthodes de mise en correspondance de pixels. Nous détaillons plus particulièrement les méthodes par propagation de germes

1. <http://vision.middlebury.edu/stereo/>

et les méthodes fondées sur les régions.

- **Chapitre 3** – La mise en correspondance par propagation de germes est une alternative intéressante à la mise en correspondance locale « classique ». La zone de recherche des correspondants est réduite aux voisinages d'appariements fiables appelés *germes*, en faisant l'hypothèse que, généralement, deux pixels voisins ont des correspondants proches. Cela permet de réduire les temps de calculs et d'éviter certaines ambiguïtés. Cependant, le succès de ce type de méthode est fortement dépendant du choix de ces germes. Nous proposons une étude de l'étape de sélection des germes. Nous nous concentrons plus particulièrement sur la mise en correspondance de points d'intérêt. Il s'agit de points de l'image ayant des caractéristiques intéressantes pour une application donnée. Dans notre cas, nous avons besoin de pixels qui peuvent être mis en correspondance de manière sûre. Nous présentons également dans ce chapitre un état de l'art de la détection de points d'intérêt.
- **Chapitre 4** – Une première contribution de ce travail porte sur l'étude de la sélection de germes par appariement de points d'intérêt. Nous comparons quatorze détecteurs bien connus associés à cinq mesures de corrélation. Certaines de ces mesures sont conçues pour être robustes à l'un des principaux problèmes de la mise en correspondance stéréoscopique binoculaire : les ruptures de profondeur. En outre, cette étude donne des conseils sur la manière de choisir les paramètres des différentes méthodes afin de trouver les meilleurs germes possibles selon des critères donnés. Ensuite, ces germes sont utilisés avec deux approches de propagation qui sont évaluées, analysées et comparées.
- **Chapitre 5** – Une deuxième contribution porte sur une amélioration de la mise en correspondance par propagation de germes. Il s'agit d'une technique multi-mesure qui permet d'obtenir un meilleur compromis entre le taux d'appariements corrects et leur répartition dans l'image qu'avec une seule mesure. Le principe de cette technique consiste à commencer par propager les germes dans les zones qui ne présentent pas de difficultés particulières avec une mesure de corrélation « classique », puis d'utiliser une mesure robuste plus performante dans les zones difficiles.
- **Chapitre 6** – Une troisième contribution porte sur une nouvelle approche de mise en correspondance dense fondée sur les régions. Différentes segmentations couleur sont utilisées. Ensuite, plusieurs instances d'un modèle de surface sont calculées pour les différentes régions à partir de correspondances initiales tirées au sort. Pour chaque pixel, chaque instance donne un pixel correspondant qui est considérée comme un vote. Finalement, le pixel qui reçoit le plus de voix est sélectionné comme correspondant final. Cette approche est relativement simple à implémenter et s'avère efficace en donnant des résultats compétitifs vis-à-vis de l'état de l'art.

2 Mise en correspondance de pixels pour la stéréovision binoculaire

LE problème de la mise en correspondance (ou appariement) de pixels consiste à retrouver, sur deux images d'une même scène prises depuis des positions différentes, les deux pixels qui correspondent à la projection d'un même élément de la scène, voir figure 1.2, page 23. Durant ces vingt dernières années, ce problème a été largement étudié et de nombreuses méthodes d'appariement ont été proposées. Elles peuvent être regroupées en différentes familles. Dans ce travail, nous avons choisi de les regrouper comme illustré par la figure 2.1. Deux familles se dégagent : les méthodes *locales* et les méthodes *globales*. Parmi les méthodes locales, nous pouvons distinguer une technique particulière de propagation de germes. Récemment, des méthodes mixtes et fondées sur une segmentation de l'image en régions sont apparues. Ces dernières méthodes se retrouvent en tête du classement proposé au sein de l'université de Middlebury¹ qui évalue et compare les méthodes de mise en correspondance dense [Scharstein 02].

Dans ce mémoire, nous traiterons plus particulièrement les méthodes suivantes :

- **Méthodes locales classiques** – Ces méthodes recherchent entre les deux images les pixels dont les voisinages se ressemblent le plus en utilisant des mesures de corrélation. Pour chaque pixel d'une image du couple, la recherche du correspondant dans l'autre image se fait parmi un ensemble de pixels candidats qui constituent une zone de recherche. Sans connaissance *a priori*, cette zone peut s'étendre à l'ensemble des pixels de l'autre image, mais il est possible de la réduire, en particulier en utilisant le modèle géométrique du capteur, voir figure 1.2, page 23.
- **Méthodes locales par propagation de germes** – Il s'agit d'un cas particulier de

1. <http://vision.middlebury.edu/stereo/>

méthodes locales où la zone de recherche des correspondants est réduite au voisinage d'appariements fiables, appelés *germes*, en faisant l'hypothèse que, généralement, deux pixels voisins ont des correspondants proches. Les appariements ainsi calculés servent à leur tour de germes, et le processus est réitéré jusqu'à l'obtention d'un résultat dense ou quasi-dense. Nous pouvons distinguer deux approches de propagation :

- ▶ *Approche séquentielle* – Un seul germe est propagé à chaque itération. Il est sélectionné selon un critère prédéfini.
- ▶ *Approche simultanée* – Un ensemble de germes est propagé à chaque itération.
- **Méthodes globales** – Ces méthodes ont pour objectif de minimiser par une technique d'optimisation mathématique, une fonction de coût représentant l'erreur induite par les correspondances trouvées.
- **Méthodes fondées sur les régions** – Elles utilisent une segmentation de l'image en régions de couleur homogène et calculent les paramètres d'un modèle de surface pour chaque région. Une initialisation par une méthode locale est généralement effectuée et une technique d'optimisation globale permet de raffiner les résultats.

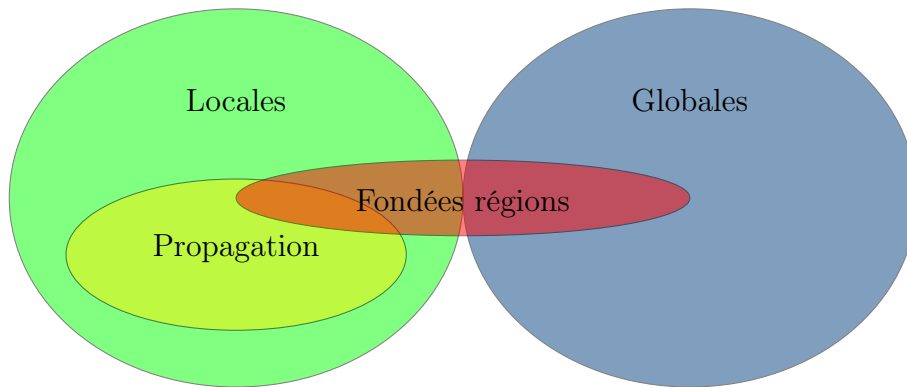


FIGURE 2.1: Familles de méthodes de mise en correspondance de pixels.

2.1 Notations

On note $\mathbf{p}_{i,j}^g$ le pixel situé sur la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne de l'image g . On note g l'image gauche et d l'image droite. Les coordonnées du pixel considéré sont telles que $i \in [0; h[$

et $j \in [0; w[$, où h est le nombre de lignes de l'image et w est le nombre de colonnes. On suppose que les deux images d'un couple stéréoscopique ont les mêmes dimensions.

- **Fonction image** I^g – La valeur d'un pixel est donnée par la « *fonction image discrétisée* » :

$$I^g : [0; h[\times [0; w[\rightarrow \mathbb{N}^c$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \mapsto I^g \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \quad (2.1)$$

où $I^g(\begin{bmatrix} i \\ j \end{bmatrix}^\top)$ est la valeur de dimension c du pixel $\mathbf{p}_{i,j}^g$. On a donc, par exemple, $c = 1$ si on considère les niveaux de gris ou $c = 3$ si on considère un système de couleur. Pour simplifier, on prendra $c = 1$ dans la suite de ce mémoire.

- **Vecteur de disparité** $\mathbf{d}_{i,j}^{g \rightarrow d}$ – Le résultat d'une mise en correspondance peut être représenté par un *vecteur de disparité* $\mathbf{d}_{i,j}^{g \rightarrow d}$ qui décrit la position relative de deux pixels correspondants entre deux images g et d , voir figure 2.2. L'image de départ, ici g , est l'image de référence. Dans la suite de ce mémoire, sauf indication contraire, nous prenons l'image gauche comme image de référence. Pour simplifier les notations, on note alors $\mathbf{d}_{i,j}^{g \rightarrow d} = \mathbf{d}_{i,j}$. Le vecteur de disparité associé au pixel $\mathbf{p}_{i,j}^g$ est tel que $\mathbf{d}_{i,j} \in ([d_{i_{\min}}; d_{i_{\max}}] \times [d_{j_{\min}}; d_{j_{\max}}]) \cup \{\mathbf{occ}\}$. Si $\mathbf{d}_{i,j} \neq \mathbf{occ}$ alors :

$$\begin{bmatrix} i' \\ j' \end{bmatrix} = \begin{bmatrix} i \\ j \end{bmatrix} + \mathbf{d}_{i,j} \quad (2.2)$$

où $\mathbf{p}_{i',j'}^d$ est le pixel de l'image droite correspondant au pixel $\mathbf{p}_{i,j}^g$, où $d_{i_{\min}}, d_{i_{\max}}, d_{j_{\min}}$ et $d_{j_{\max}}$ sont les bornes des intervalles des valeurs possibles pour les composantes du vecteur disparité et où \mathbf{occ} est un vecteur de disparité symbolique attribué aux pixels dont les correspondants sont occultés.

Comme nous allons le voir dans la suite de ce document, il est possible de rectifier les images de telle sorte que le correspondant d'un pixel de l'image gauche se situe sur la même ligne dans l'image de droite, voir § 2.2.2. Ainsi, la première composante du vecteur de disparité est toujours nulle. Dans ce cas, le vecteur de disparité peut être remplacé par un scalaire $d_{i,j}$ appelé *disparité* ou *valeur de disparité* tel que :

$$\mathbf{d}_{i,j} = \begin{bmatrix} 0 \\ d_{i,j} \end{bmatrix}. \quad (2.3)$$

Dans la suite de ce mémoire, sauf indication contraire, nous nous plaçons dans ce cas.

- **Fonction de disparité** d – Le champ des vecteurs de disparité d'un couple d'images peut être représenté par une *fonction de disparité* :

$$d : [0; h[\times [0; w[\rightarrow ([d_{i_{\min}}; d_{i_{\max}}] \times [d_{j_{\min}}; d_{j_{\max}}]) \cup \{\mathbf{occ}\}$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \mapsto \mathbf{d}_{i,j}. \quad (2.4)$$

■ 2 Mise en correspondance de pixels pour la stéréovision binoculaire

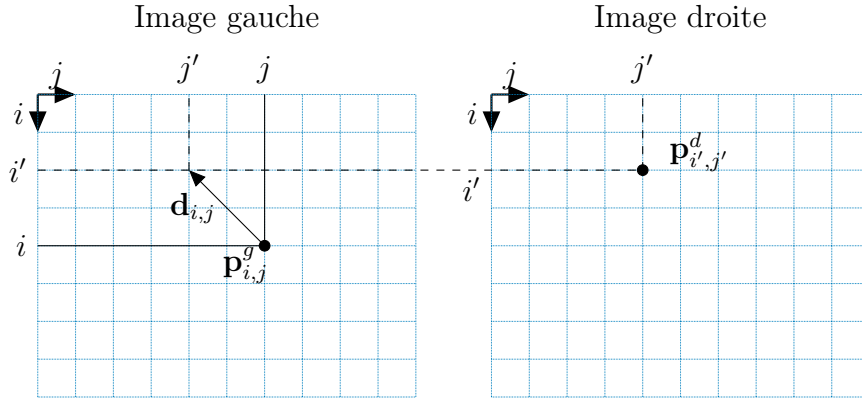


FIGURE 2.2: *Vecteur de disparité.* – Il est tel que $\mathbf{d}_{i,j} = [i' - i \quad j' - j]^\top$.

- **Fonction de correspondance m** – Nous pouvons également définir une *fonction de correspondance* qui donne les coordonnées des pixels correspondants :

$$m : [0; h[\times [0; w[\rightarrow ([0; h[\times [0; w[) \cup \emptyset$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \mapsto \begin{bmatrix} i' \\ j' \end{bmatrix} . \quad (2.5)$$

- **Carte de disparité** – Enfin, le résultat d’une mise en correspondance de pixels peut être visualisé par une carte de disparité. Il s’agit d’une image dont le niveau de gris de chaque pixel est proportionnel à sa disparité dans l’image de référence. Plus il est clair plus l’élément qui se projette sur ce pixel est proche du capteur, voir figure 2.3.



FIGURE 2.3: *Couple d’images et carte de disparité.* Sur la carte de disparité, les pixels noirs sont les pixels dont le correspondant est occulté ou inconnu.

2.2 Méthodes locales

Les méthodes locales mesurent la similarité des voisinages de deux pixels. Ce voisinage est appelé *zone d'agrégation* ou *fenêtre de corrélation*. Pour simplifier, on suppose ce voisinage carré $(n \times n = (2r + 1) \times (2r + 1))$, où $r \in \mathbb{N}^*$ et centré sur le pixel considéré.

On note \mathbf{a} le vecteur de dimension n^2 contenant les valeurs des pixels voisins du pixel considéré sur l'image de référence et \mathbf{b} celui de l'autre image. Les méthodes locales supposent que les voisinages de deux correspondants se ressemblent, c'est-à-dire :

$$\mathbf{a} \approx \mathbf{b}. \quad (2.6)$$

Cette similarité est évaluée entre chaque pixel de l'image de référence et chaque pixel de sa *zone de recherche* dans l'autre image. Sur la figure 2.4, le pixel de l'image gauche à mettre en correspondance est marqué d'un point noir. Son correspondant est recherché parmi les pixels situés dans la zone de recherche. Dans cet exemple, la zone de recherche considérée est monodimensionnelle. Elle est définie en fonction de la position du pixel de gauche dont on cherche le correspondant. Pour chaque candidat, un score de corrélation qui prend en compte le voisinage (fenêtres de corrélation en gris de dimensions 3×3 ici) est calculé. Enfin, le candidat qui donne le meilleur score est retenu comme correspondant (point noir). REMARQUE : les mesures de corrélation peuvent quantifier une similarité ou bien une dissemblance. Dans la suite de ce document, nous considérons que le « meilleur » score correspond donc au maximum d'une mesure de similarité ou au minimum d'une mesure de dissemblance.

Plus formellement, on définit une mesure de corrélation générique c par :

$$\begin{aligned} c : \mathbb{N}^{n^2}, \mathbb{N}^{n^2} &\rightarrow \mathbb{N} \\ \mathbf{a}, \mathbf{b} &\mapsto c(\mathbf{a}, \mathbf{b}). \end{aligned} \quad (2.7)$$

2.2.1 Mise en correspondance locale classique : algorithme élémentaire

Le candidat de la zone de recherche dont le voisinage ressemble le plus à celui du pixel considéré, c'est-à-dire celui qui obtient le meilleur score de corrélation est sélectionné comme correspondant. Cette stratégie est nommée stratégie du *winner-take-all* (*wta*). L'algorithme 2.1 résume les différentes étapes de la mise en correspondance locale classique.

Les erreurs sont fréquentes avec l'algorithme élémentaire, en particulier dans les zones occultées ou proches des discontinuités de profondeur (nous avons vu à la section 1.2, page 22, que l'hypothèse de l'expression (2.6) n'est pas toujours vérifiée).

Pour limiter ces erreurs, il existe deux possibilités :

■ 2 Mise en correspondance de pixels pour la stéréovision binoculaire

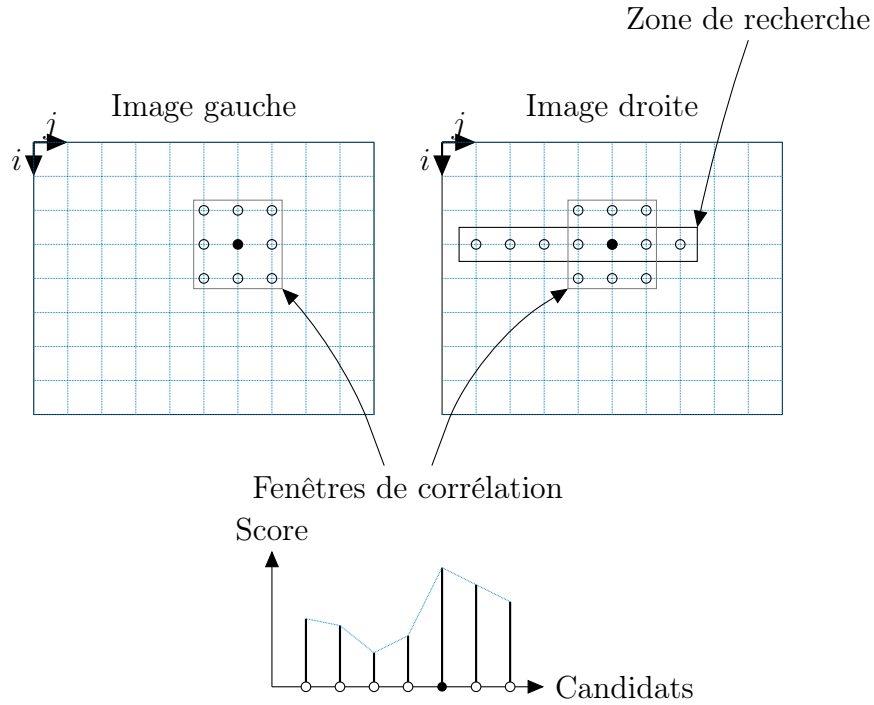


FIGURE 2.4: *Mise en correspondance locale.*

```

1:  $s_{\max} \leftarrow -\infty$ 
2: pour chaque pixel  $\mathbf{p}_{i,j}^g$  de l'image gauche faire
3:   pour chaque candidat  $\mathbf{p}_{i',j'}^d$  de la zone de recherche de l'image droite faire
4:      $s \leftarrow c(\mathbf{a}, \mathbf{b})$ , score de corrélation entre les valeurs des pixels aux voisinages de  $\mathbf{p}_{i,j}^g$ 
      (stockées dans le vecteur  $\mathbf{a}$ ), et de,  $\mathbf{p}_{i',j'}^d$  (stockées dans le vecteur  $\mathbf{b}$ )
5:     si  $s > s_{\max}$  alors
6:        $s_{\max} \leftarrow s$ 
7:        $\hat{i}' \leftarrow i'$ 
8:        $\hat{j}' \leftarrow j'$ 
9:     fin si
10:  fin pour
11:   $m([i \ j]) \leftarrow [\hat{i}' \ \hat{j}']$ 
12: fin pour

```

ALGORITHME 2.1: *Algorithme élémentaire de mise en correspondance locale classique.*

- **ajout de contraintes** : elles permettent d'éviter des appariements qui ne vérifient pas certaines propriétés ;
- **prise en compte des difficultés de la mise en correspondance** : il s'agit d'utiliser certaines techniques qui permettent de prendre en compte certaines particularités de l'image, comme, par exemple, la présence de contours.

2.2.2 Ajout de contraintes

Nous distinguons les contraintes *a priori* et *a posteriori* :

- **contrainte *a priori*** : l'objectif est de limiter le risque d'effectuer un appariement erroné en éliminant des candidats et donc en réduisant la zone de recherche à l'aide de la géométrie épipolaire ; cette contrainte est également appelée *contrainte épipolaire* ;
- **contraintes *a posteriori*** : après avoir effectué la mise en correspondance, l'objectif est d'éliminer les appariements les moins fiables qui ne vérifient pas certaines propriétés.

2.2.2.1 Contrainte *a priori* : la contrainte épipolaire

Des contraintes géométriques permettent de réduire l'ensemble des correspondances possibles, en particulier en utilisant la géométrie épipolaire qui décrit les relations entre deux vues. Il est possible de réduire la zone de recherche à une seule droite en utilisant des propriétés de la géométrie épipolaire comme illustré sur la figure 2.5. Soit \mathbf{P} un point de la scène. Le plan qui passe par \mathbf{P} et les centres optiques \mathbf{O}^g et \mathbf{O}^d des deux caméras est appelé le plan épipolaire. Il coupe les plans images gauche et droit suivant deux droites appelées droites épipolaires conjuguées. Les correspondants des pixels situés sur la droite épipolaire gauche se trouvent sur la droite épipolaire conjuguée droite. Les paramètres de la géométrie épipolaire sont exprimés par une matrice de taille 3×3 appelée *matrice fondamentale* [Hartley 03].

La *rectification épipolaire* est une transformation géométrique appliquée aux deux images de telle sorte que les droites épipolaires soient toutes parallèles et horizontales dans les deux images, voir figure 2.6. Cela permet de réduire la recherche du correspondant d'un pixel situé sur la ligne i et la colonne j de l'image gauche aux pixels de la même ligne sur l'autre image situés à gauche de la colonne j . La rectification épipolaire passe généralement par l'estimation de la matrice fondamentale. Cette estimation est généralement réalisée à partir de correspondance de points d'intérêt (voir chapitre 3 pour la détection de ces points d'intérêt). Le lecteur intéressé par différentes techniques d'estimation de la matrice fondamentale peut consulter [Hartley 03, Bartoli 03].

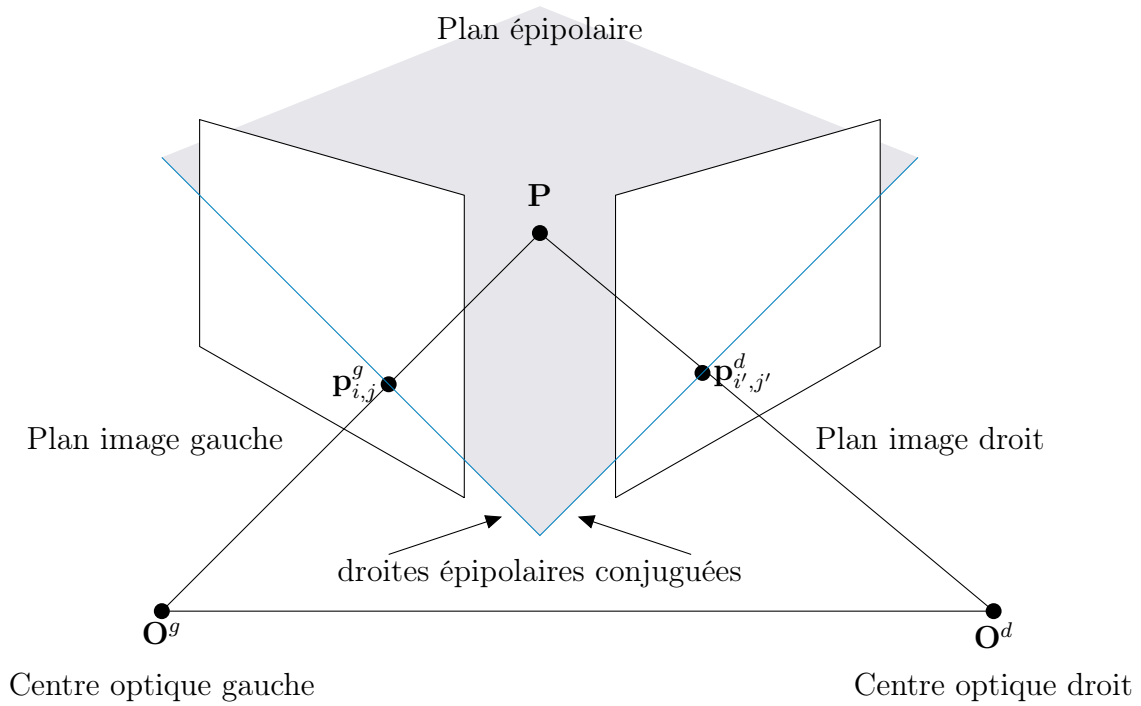


FIGURE 2.5: *Géométrie épipolaire.*

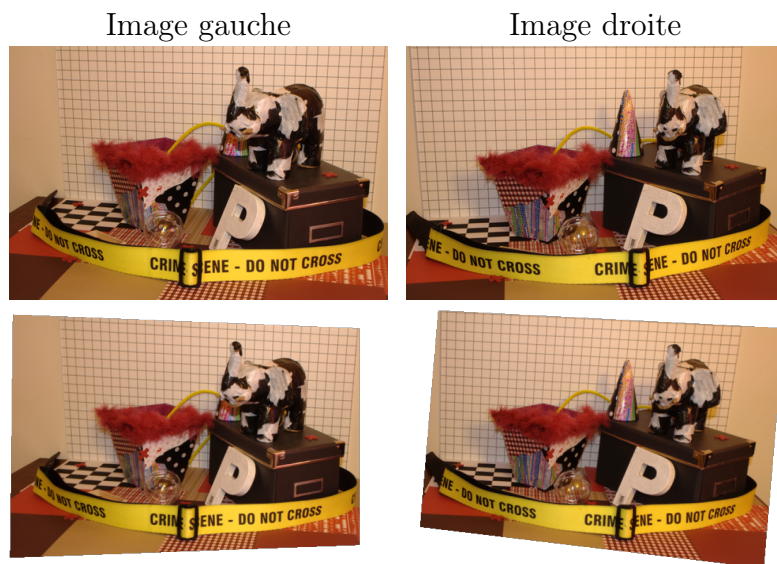


FIGURE 2.6: *Rectification épipolaire.* – La première ligne présente un couple d'images et sa version rectifiée est sur la seconde ligne.

2.2.2.2 Contraintes *a posteriori*

Les contraintes *a posteriori* sont utilisées après la mise en correspondance pour éliminer les appariements qui ne sont pas suffisamment fiables. Toutefois, ces contraintes peuvent également éliminer de bons appariements, réduisant ainsi la densité du résultat obtenu. Il faut donc réaliser un compromis. Ces contraintes de mise en correspondance sont nombreuses et nous ne présentons ici que les plus couramment utilisées.

- **Contrainte sur le score de corrélation** – Il s'agit de la contrainte la plus utilisée. Elle consiste à éliminer les appariements dont le score de corrélation est inférieur à un seuil, c'est-à-dire à ne conserver que les appariements $(\mathbf{p}_{i,j}^g, \mathbf{p}_{i,j'}^d)$ tels que :

$$c(\mathbf{a}, \mathbf{b}) > t_s \quad (2.8)$$

où \mathbf{a} est le vecteur des valeurs des pixels voisins de $\mathbf{p}_{i,j}^g$, \mathbf{b} est le vecteur des valeurs des pixels voisins de $\mathbf{p}_{i,j'}^d$ et $t_s \in \mathbb{R}$ est un seuil.

REMARQUE : certaines mesures de corrélation sont normalisées et donnent une valeur comprise entre c_{\min} et c_{\max} . Dans ce cas, le choix du seuil peut être plus intuitif que si l'on ne connaît pas ces bornes.

- **Contrainte d'ordre** – L'ordre des pixels le long de la droite épipolaire doit être le même que celui de leurs correspondants le long de la droite épipolaire conjuguée, voir § 2.2.2.1 et figure 2.7. Cette contrainte peut se traduire par :

$$\begin{aligned} \left(m \begin{bmatrix} i & j_1 \end{bmatrix}^\top = \begin{bmatrix} i & j'_1 \end{bmatrix}^\top \wedge \right. \\ \left. m \begin{bmatrix} i & j_2 \end{bmatrix}^\top = \begin{bmatrix} i & j'_2 \end{bmatrix}^\top \wedge \right. \\ \left. j_1 < j_2 \right) \implies j'_1 < j'_2. \end{aligned} \quad (2.9)$$

- **Contrainte d'unicité** – Cette contrainte stipule qu'un pixel d'une image peut être apparié avec au plus un pixel de l'autre image [Ayache 89], voir figure 2.8. La contrainte d'unicité est vérifiée quand :

$$\begin{aligned} \left(m^{g \rightarrow d} \begin{bmatrix} i & j_1 \end{bmatrix}^\top = \begin{bmatrix} i & j'_1 \end{bmatrix}^\top \implies \right. \\ \left. \forall j_2 \neq j_1, \quad m^{g \rightarrow d} \begin{bmatrix} i & j_2 \end{bmatrix}^\top \neq \begin{bmatrix} i & j'_1 \end{bmatrix}^\top \right) \\ \wedge \left(m^{d \rightarrow g} \begin{bmatrix} i & j'_3 \end{bmatrix}^\top = \begin{bmatrix} i & j_3 \end{bmatrix}^\top \implies \right. \\ \left. \forall j'_4 \neq j'_3, \quad m^{d \rightarrow g} \begin{bmatrix} i & j'_4 \end{bmatrix}^\top \neq \begin{bmatrix} i & j_3 \end{bmatrix}^\top \right). \end{aligned} \quad (2.10)$$

- **Contrainte de symétrie** – La contrainte de symétrie introduite dans [Fua 93], consiste à vérifier qu'on retrouve les mêmes appariements quelle que soit l'image de référence, voir figure 2.9 :

$$\begin{aligned} m^{g \rightarrow d} \begin{bmatrix} i & j \end{bmatrix}^\top = \begin{bmatrix} i & j' \end{bmatrix}^\top \Leftrightarrow \\ m^{d \rightarrow g} \begin{bmatrix} i & j' \end{bmatrix}^\top = \begin{bmatrix} i & j \end{bmatrix}^\top. \end{aligned} \quad (2.11)$$

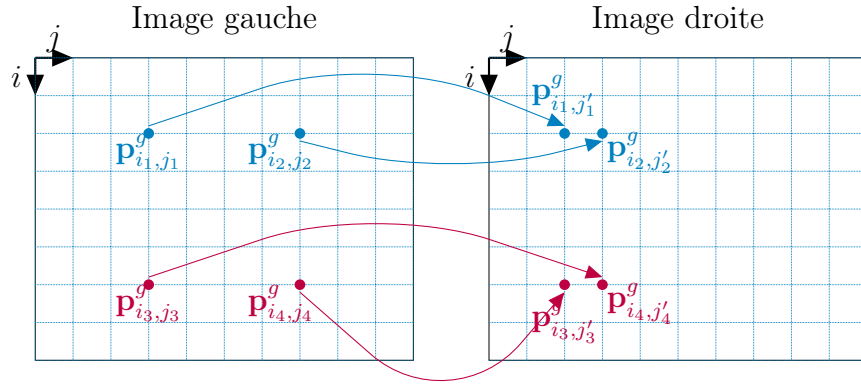


FIGURE 2.7: *Contrainte d'ordre.* – En bleu (en haut), on montre un exemple en configuration parallèle où la contrainte d'ordre est vérifiée, en violet (en bas) un exemple où elle ne l'est pas.

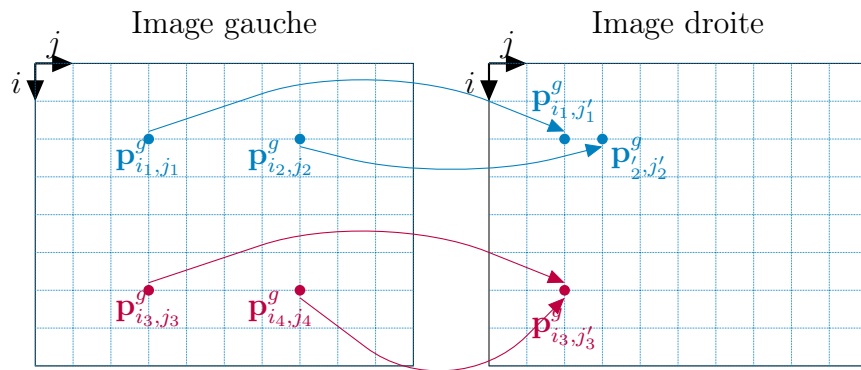


FIGURE 2.8: *Contrainte d'unicité.* – En bleu (en haut), on montre un exemple où la contrainte d'unicité est vérifiée, en violet (en bas) un exemple où elle ne l'est pas.

Cette contrainte peut être imposée en effectuant une *vérification bidirectionnelle*, c'est-à-dire en effectuant une mise en correspondance de l'image gauche vers l'image droite, une mise en correspondance de l'image droite vers l'image gauche et en vérifiant que l'on retrouve bien les mêmes appariements dans les deux sens.

- **Contrainte de symétrie faible** – La contrainte précédente est très forte car sa vérification peut éliminer beaucoup d'appariements corrects. Une variante plus tolérante, mise en œuvre par la *vérification bidirectionnelle souple* introduite dans [Crouzil 97], et qui utilise un seuil $t_s \in \mathbb{R}$ est donnée par :

$$\begin{aligned} m^{g \rightarrow d}([i \ j]^\top) &= [i \ j']^\top \wedge \\ m^{d \rightarrow g}([i \ j']^\top) &= [i \ j_2]^\top \implies \\ |j - j_2| &< t_s. \end{aligned} \quad (2.12)$$

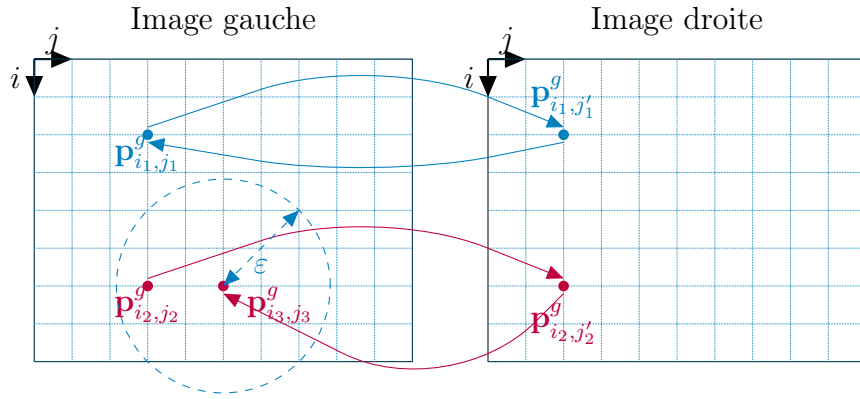


FIGURE 2.9: *Contrainte de symétrie et de symétrie faible.* – En bleu (en haut), on montre un exemple où la contrainte de symétrie est vérifiée, en violet (en bas), un exemple où elle ne l'est pas. Cependant, sur cet exemple, la contrainte de symétrie faible est vérifiée si on prend $t_s = \varepsilon$.

- **Consistance faible** – La consistance faible est une alternative à la contrainte de symétrie présentée dans [Gong 03]. Si un point de la scène se projette en $\mathbf{p}_{i,j}^g$ et $\mathbf{p}_{i,j'}^d$, alors le correspondant de $\mathbf{p}_{i,j}^g$ est $\mathbf{p}_{i,j'}^d$, sauf s'il est occulté. Dans ce cas, il ne peut être occulté que par un point plus proche de la caméra et ayant donc une disparité plus grande, ce qui se traduit par :

$$\text{Si } j + d_{i,j} = j_2 \text{ alors pour tout } j' > j, j' + d_{i,j'} \neq j_2 \text{ et pour tout } j'_2 < j_2, j'_2 + d_{i,j'_2} \neq j \quad (2.13)$$

- **Limite du gradient de disparité** – Cette contrainte impose que les disparités ne varient pas brusquement entre deux pixels voisins. Elle est définie dans [Horaud 93] par :

$$\begin{aligned} m([i_1 \ j_1]^\top) &= [i_1 \ j'_1]^\top \wedge \\ m([i_2 \ j_2]^\top) &= [i_2 \ j'_2]^\top \implies \\ 2 \frac{\|([i_1 \ j'_1]^\top - [i_2 \ j'_2]^\top) - ([i_1 \ j_1]^\top - [i_2 \ j_2]^\top)\|}{\|([i_1 \ j'_1]^\top - [i_2 \ j'_2]^\top) + ([i_1 \ j_1]^\top - [i_2 \ j_2]^\top)\|} &< t_g \end{aligned} \quad (2.14)$$

où \mathbf{p}_{i_1, j_1}^g et \mathbf{p}_{i_2, j_2}^g sont des pixels qui sont les projections de deux points de la scène qui appartiennent à la même surface et où $t_g \in \mathbb{R}$ est un seuil.

- **Contrainte de rang** – Cette contrainte, présentée dans [Banks 01], compare la valeur du pixel central avec celle des autres pixels de la fenêtre de corrélation et impose que les voisins dont les valeurs sont supérieures à celles du pixel central soient les mêmes pour les deux pixels correspondants. Sur l'exemple donné par la figure 2.10, sur la première ligne, on montre les valeurs contenues dans une fenêtre de corrélation 3×3 autour du pixel considéré à gauche en (a) et autour de deux candidats à droite en (b) et (c). Sur la seconde ligne, on montre, pour chacune des fenêtres, les différences entre la valeur du pixel central et celle de chacun de ses voisins. On constate alors que sur cet exemple, ces différences pour le candidat (b) ont toutes le même signe que celles du pixel considéré (a) : la contrainte de rang est respectée. En revanche, ce n'est pas le cas pour toutes les différences du candidat (c) (en gras) : la contrainte de rang n'est pas respectée. Cette contrainte peut s'écrire par :

$$\begin{aligned} m([i \ j]^\top) &= [i \ j']^\top \implies \\ \forall (k, l) \in \left[-\frac{n}{2}; \frac{n}{2}\right]^2, \quad (I_{i+k, j+l}^g - I_{i, j}^g) (I_{i+k, j'+l}^d - I_{i, j'}^d) &> 0. \end{aligned} \quad (2.15)$$

- **Contrainte de fiabilité** – Un seuillage sur des mesures de fiabilité permet d'éliminer les appariements les moins fiables. Soit r une fonction qui mesure la fiabilité à partir des m scores de corrélation entre le pixel considéré $\mathbf{p}_{i, j}^g$ et les m candidats \mathbf{p}_{i_k, j'_k}^d de la zone de recherche avec $k \in [0; m[$:

$$\begin{aligned} r : \mathbb{R}^m &\rightarrow \mathbb{R} \\ \begin{bmatrix} \vdots \\ c(\mathbf{a}, \mathbf{b})_k \\ \vdots \end{bmatrix} &\mapsto r \left(\begin{bmatrix} \vdots \\ c(\mathbf{a}, \mathbf{b})_k \\ \vdots \end{bmatrix} \right). \end{aligned} \quad (2.16)$$

Notons $\mathbf{c} = [\dots \ c(\mathbf{a}, \mathbf{b})_k \ \dots]^\top$. La contrainte peut alors s'écrire :

$$r(\mathbf{c}) < t_f \quad (2.17)$$

	Gauche			Droite		
	(a)			(b)	(c)	
Niveaux de gris	250	200	198	255	205	203
	100	104	110	105	109	115
	50	48	78	55	53	93
Différences	146	96	94	146	96	94
	-4	0	6	-4	0	6
	-54	-56	-26	-54	-56	-16

FIGURE 2.10: *Contrainte de rang.*

où $t_f \in \mathbb{R}$ est un seuil. Nous pouvons distinguer les mesures de fiabilité (fonctions r) suivantes :

- *Mesure d'ambiguïté* – Elle mesure la distance en pixels entre le candidat qui a obtenu le meilleur score de corrélation et celui qui a le second meilleur score, voir figure 2.11. Soit :

$$\text{amb}(\mathbf{c}) = |(\mathbf{c})_{1:m} - (\mathbf{c})_{2:m}| \quad (2.18)$$

où $(\mathbf{c})_{1:m}$ et $(\mathbf{c})_{2:m}$ sont les deux premiers éléments du vecteur \mathbf{c} dont les éléments sont triés dans l'ordre décroissant. La mesure d'ambiguïté n'est calculée que si le second plus grand score de corrélation est proche du premier. Soit $d_s = |(\mathbf{c})_{1:m} - (\mathbf{c})_{2:m}|$. La mesure d'ambiguïté est donnée par :

$$\text{AMB}(\mathbf{c}) = \begin{cases} \text{amb}(\mathbf{c}) & \text{si } d_s < t_a \\ 0 & \text{sinon} \end{cases} \quad (2.19)$$

où $t_a \in \mathbb{R}$ est un seuil.

- *Mesure d'imprécision* – Elle est donnée par la distance entre les pixels k_1 et k_2 qui définissent l'intervalle dans lequel les scores sont proches du meilleur score, c'est-à-dire très élevés, voir figure 2.12 :

$$\text{IMP}(\mathbf{c}) = |k_1 - k_2|, \text{ où } k_1 \text{ et } k_2 \text{ sont tels que } \forall k \in [k_1; k_2], c_k > t_i \quad (2.20)$$

où $t_i \in \mathbb{R}$ est un seuil.

2.2.3 Prise en compte des difficultés de la mise en correspondance

Nous distinguons deux façons de prendre en compte les principales difficultés de la mise en correspondance :

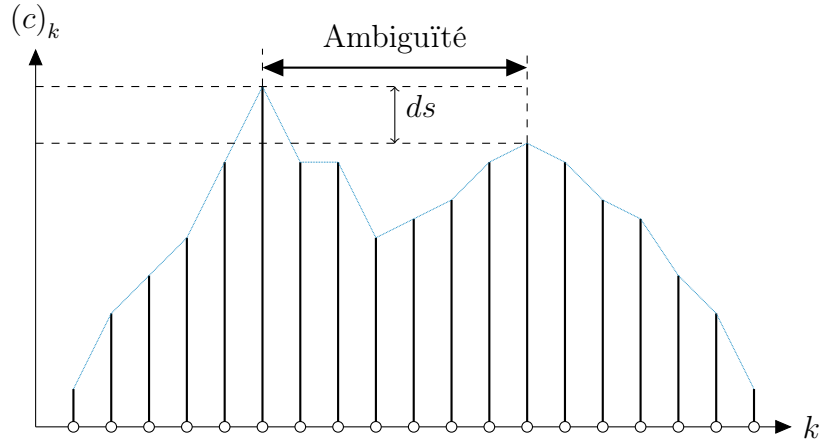


FIGURE 2.11: *Mesure d'ambiguïté.* – Les candidats de la zone de recherche sont marqués en abscisse. La mesure d'ambiguïté correspond à la distance (en pixels) entre les deux candidats qui obtiennent les deux meilleurs scores de corrélation (en ordonnée).

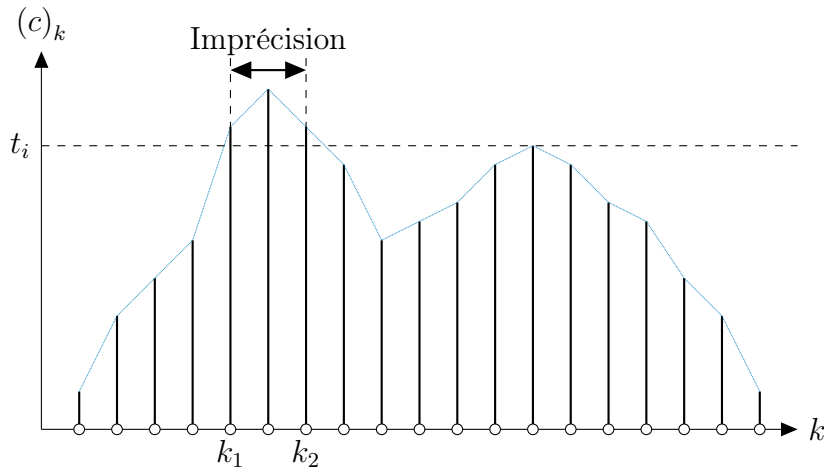


FIGURE 2.12: *Mesure d'imprécision.* – Les candidats de la zone de recherche sont marqués en abscisse, la mesure d'imprécision correspond au nombre de pixels dont le score de corrélation, en ordonnée, est supérieur à un seuil t_i .

- **mesure de corrélation adaptée au contexte** : certaines mesures de corrélation ont été conçues pour être robustes à certaines difficultés, par exemple aux zones proches des discontinuités de profondeur ;
- **fenêtre de corrélation adaptée au contexte** : le but est d'adapter la morphologie de la fenêtre de corrélation afin de ne prendre en compte que la partie du voisinage la plus pertinente (la totalité des pixels appartenant aux voisinages carrés de deux correspondants ne se ressemblent pas forcément près des occultations et des discontinuités de profondeur, voir § 1.2, page 22).

2.2.3.1 Mesures de corrélation

Plusieurs familles de mesures de corrélation peuvent être distinguées [Chambon 11] :

- **Famille classique** – Elle regroupe les mesures qui font appel aux statistiques classiques de la distribution des différences des valeurs des pixels entre les deux fenêtres de corrélation. Il s'agit par exemple de mesures calculant une norme ou une variance.
- **Famille croisée** – Elle contient les mesures qui utilisent une corrélation croisée comme le coefficient de corrélation linéaire.
- **Famille dérivée** – Elle comprend les mesures qui s'appliquent aux variations des valeurs des pixels.
- **Famille non paramétrique** – Les mesures de cette famille s'intéressent à l'ordre des niveaux de gris en se fondant sur des statistiques non paramétriques.
- **Famille robuste** – Elle regroupe les mesures qui se basent sur des outils de statistiques robustes pour ne pas prendre en compte les valeurs des pixels de la fenêtre de corrélation dans l'image de référence qui n'ont pas de correspondant dans la fenêtre de corrélation du pixel candidat de l'autre image (en cas notamment d'occultation ou de discontinuité de profondeur).

Nous présentons ci-dessous, pour chaque famille, les mesures de corrélation qui donnent les meilleurs résultats selon l'évaluation réalisée dans [Chambon 11] ainsi que les mesures qui seront utilisées dans la suite de ce document. Une taxonomie détaillée des mesures de corrélation pour la mise en correspondance de pixels peut également être trouvée dans ce même document.

- **Norme L_P** (famille classique) – Il s'agit d'une mesure de distance :

$$D_P(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_P^P. \quad (2.21)$$

D_1 est la somme des valeurs absolues des différences (SAD : *Sum of Absolute Differences*) et D_2 la somme des carrés des différences (SSD : *Sum of Squared Differences*).

- **Somme des valeurs absolues des différences centrées (ZSAD pour Zero mean Sum of Absolute Differences)** (famille classique) – Un cas particulier de distance est :

$$\text{ZSAD}(\mathbf{a}, \mathbf{b}) = \|(\mathbf{a} - \boldsymbol{\mu}(\mathbf{a})) - (\mathbf{b} - \boldsymbol{\mu}(\mathbf{b}))\|_1 \quad (2.22)$$

où $\mu(\mathbf{a})$ est le vecteur de dimension n^2 contenant la moyenne des éléments du vecteur \mathbf{a} . Cette mesure centre les valeurs des fenêtres de corrélation afin d'être invariante au *biais*, c'est-à-dire $\text{ZSAD}(\mathbf{a} + x, \mathbf{b} + y) = \text{ZSAD}(\mathbf{a}, \mathbf{b})$ avec $(x, y) \in \mathbb{R}^2$.

- **Corrélation croisée normalisée (NCC pour *Normalized Cross Correlation*)** (famille croisée) – La corrélation croisée peut également être employée comme mesure de similarité :

$$\text{NCC}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \quad (2.23)$$

où \cdot représente le produit scalaire et où les valeurs des fenêtres de corrélation sont normalisées afin que la mesure soit invariante au *gain*, c'est-à-dire $\text{NCC}(\mathbf{a}x, \mathbf{b}y) = \text{NCC}(\mathbf{a}, \mathbf{b})$ avec $(x, y) \in \mathbb{R}^{*2}$.

- **Corrélation croisée centrée normalisée (ZNCC pour *Zero mean Normalized Cross Correlation*)** (famille croisée) – Il existe également une version centrée de NCC, le coefficient de corrélation linéaire :

$$\text{ZNCC}(\mathbf{a}, \mathbf{b}) = \frac{(\mathbf{a} - \mu(\mathbf{a})) \cdot (\mathbf{b} - \mu(\mathbf{b}))}{\|\mathbf{a} - \mu(\mathbf{a})\| \|\mathbf{b} - \mu(\mathbf{b})\|} \quad (2.24)$$

Cette mesure est invariante au gain et au biais.

- **Corrélation de Moravec (MOR)** (famille croisée) – Une autre mesure de corrélation basée sur la corrélation croisée est la mesure de Moravec [Moravec 80] :

$$\text{MOR}(\mathbf{a}, \mathbf{b}) = \frac{2(\mathbf{a} - \mu(\mathbf{a})) \cdot (\mathbf{b} - \mu(\mathbf{b}))}{\|\mathbf{a} - \mu(\mathbf{a})\|^2 + \|\mathbf{b} - \mu(\mathbf{b})\|^2} \quad (2.25)$$

Cette mesure est invariante au *biais*.

- **Corrélation des champs de gradients (GC pour *Gradient Correlation*)** (famille dérivée) – Une alternative aux mesures classiques consiste à ne pas s'intéresser directement aux valeurs des pixels, mais à leurs variations au travers de leurs dérivées premières. La corrélation des gradients introduite dans [Crouzil 96] mesure la dissimilitude des vecteurs gradients contenus dans les deux voisinages et peut s'écrire :

$$\text{GC}(\mathbf{a}, \mathbf{b}) = \frac{\sum_{p=-\frac{n}{2}}^{\frac{n}{2}} \sum_{q=-\frac{n}{2}}^{\frac{n}{2}} \|\nabla I^g([i+p \ j+q]^\top) - \nabla I^d([i+p \ j'+q]^\top)\|}{\sum_{p=-\frac{n}{2}}^{\frac{n}{2}} \sum_{q=-\frac{n}{2}}^{\frac{n}{2}} \|\nabla I^g([i+p \ j+q]^\top)\| + \|\nabla I^d([i+p \ j'+q]^\top)\|} \quad (2.26)$$

où ∇I désigne le champ de vecteurs gradients. La mesure GC est invariante au biais.

- **Smooth Median Powered Deviation (SMPD_P)** (famille robuste) – Cette mesure s'appuie sur l'estimateur SMAD (*Smooth Median Absolute Deviation*) proposé dans [Rousseeuw 92]. Il s'agit de calculer une distance centrée robuste. La robustesse

est obtenue grâce au centrage par la médiane et à l'utilisation d'une distance tronquée. Ainsi, lorsqu'il existe des différences importantes entre deux voisinages de pixels correspondants dues à des occultations ou à des discontinuités de profondeur, celles-ci ne sont pas prises en compte, rendant ainsi la mesure plus robuste qu'une mesure classique (qui va prendre en compte ces grandes différences) :

$$\text{SMPD}_P(\mathbf{a}, \mathbf{b}) = \sum_{k=0}^{h-1} (|\mathbf{a} - \mathbf{b} - \text{med}(\mathbf{a} - \mathbf{b})\mathbf{1}|^P)_{k:n^2-1} \quad (2.27)$$

où $\text{med}(\mathbf{a} - \mathbf{b})$ est la médiane des différences des valeurs des deux fenêtres de corrélation et où $k : n^2 - 1$ est le $k^{\text{ème}}$ élément du vecteur des écarts à la médiane triés dans l'ordre croissant. On prend généralement $h = n^2/2$.

- **CENSUS** (famille non paramétrique) – Cette mesure, introduite dans [Zabih 94], prend en compte les différences de valeurs entre le pixel étudié et ses voisins. Soit $R^{\mathbf{a}}$ et $R^{\mathbf{b}}$ deux chaînes de bits définies de la manière suivante :

$$R^{\mathbf{a}} = \otimes_{k \in [0; n^2[} \begin{cases} 1 & \text{si } a_{\frac{n^2}{2}} > a_k \\ 0 & \text{sinon} \end{cases} \quad R^{\mathbf{b}} = \otimes_{k \in [0; n^2[} \begin{cases} 1 & \text{si } b_{\frac{n^2}{2}} > b_k \\ 0 & \text{sinon} \end{cases} \quad (2.28)$$

où \otimes est l'opérateur de concaténation et k parcourt le voisinage (de dimension n^2) de chaque pixel de la fenêtre de corrélation. Le score est donné par :

$$\text{CENSUS}(\mathbf{a}, \mathbf{b}) = d_{\text{Ham}}(R^{\mathbf{a}}, R^{\mathbf{b}}) \quad (2.29)$$

où d_{Ham} est la distance de Hamming, c'est-à-dire le nombre de bits qui diffèrent entre les deux chaînes. Cette mesure donne également de bons résultats dans les zones proches des occultations et des discontinuités de profondeur dans [Chambon 11]. Cette mesure évalue le respect de la contrainte de rang, voir § 2.2.2.2.

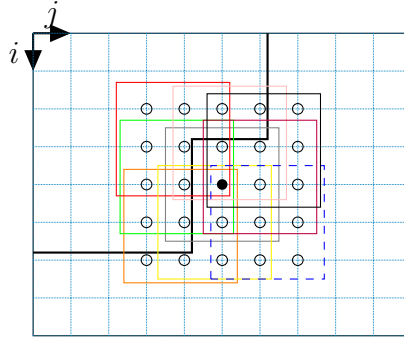
2.2.3.2 Fenêtres de corrélation

Une taille de fenêtre de corrélation trop petite ne prend pas en compte assez d'information pour être suffisamment discriminante et pour faire face à certaines difficultés, en particulier dans les zones homogènes constituées de textures répétitives. De plus, la mesure risque d'être trop sensible à de petites variations peu significatives (bruit). Inversement, une fenêtre de corrélation trop grande risque de contenir des pixels n'ayant pas la même disparité que celle du pixel considéré. Ainsi, pour deux pixels correspondants, on va obtenir des voisinages qui ne sont pas toujours similaires, voir § 1.2, page 22. Ces différences sont prises en compte par la mesure de corrélation employée et, si celle-ci n'est pas adaptée, elle donne des scores faibles pour deux correspondants et peut impliquer des erreurs d'appariements. Différentes techniques permettant d'adapter la taille ou la forme des fenêtres de corrélation pour ne prendre en compte que les pixels du voisinage qui se correspondent ont donc été proposées. Nous pouvons distinguer les méthodes suivantes :

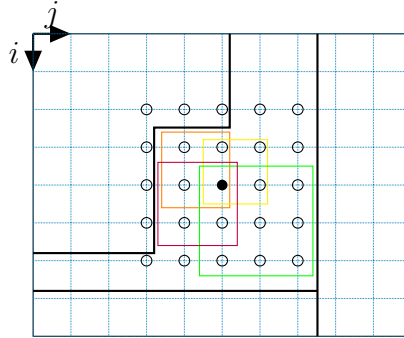
- **déformation liée des fenêtres de corrélation** : les fenêtres de corrélation ont la même forme dans l'image de référence et l'autre image ;
- **déformation indépendante des fenêtres de corrélation** : les fenêtres de corrélation n'ont pas la même forme dans l'image de référence et dans l'autre image ;
- **pondération des pixels de la fenêtre de corrélation** : un poids est attribué à chaque pixel de la fenêtre et il intervient dans le calcul du score de corrélation.

Déformation liée des fenêtres de corrélation

- **Minimisation d'un critère d'incertitude** – Dans [Kanade 94], les auteurs proposent un algorithme itératif qui estime et met à jour pour chaque pixel la valeur de disparité trouvée en adaptant à chaque itération les dimensions de la fenêtre de corrélation utilisée. Ces dimensions sont déterminées en testant plusieurs dimensions possibles et en sélectionnant celles qui minimisent un critère d'incertitude sur la disparité précédemment estimée. Le calcul de ce critère d'incertitude est basé sur un modèle statistique qui prend en compte la distribution des différences entre la disparité estimée pour le pixel considéré et celles estimées pour les autres pixels de la fenêtre de corrélation ainsi que la distribution des variations des valeurs des pixels dans la fenêtre de corrélation du candidat de l'autre image. Ainsi, la fenêtre qui minimise ce critère est utilisée pour mettre à jour la valeur de disparité estimée du pixel considéré et le processus est répété jusqu'à convergence.
- **Multi-fenêtrage fixe** – Dans [Bobick 99], neuf fenêtres de corrélation différentes sont utilisées. La première est centrée sur le pixel considéré, les huit autres sont excentrées comme illustré sur la figure 2.13 où la ligne noire épaisse indique une rupture de profondeur. Le pixel considéré est marqué d'un point noir. Les neuf fenêtres de corrélation sont marquées d'une couleur différente. Sur cet exemple, la fenêtre bleue (en pointillés) ne prend en compte que les pixels de la même disparité que le pixel considéré ; elle vérifie donc l'hypothèse de similarité (2.6) contrairement à la fenêtre grise centrée sur le pixel étudié. Le score final est donné par la fenêtre qui donne le score de similarité le plus élevé. On suppose qu'il s'agit du score donné par la fenêtre qui prend en compte le plus de pixels voisins de même disparité que le pixel considéré. Cette technique est connue sous le nom de *shiftable windows*.
- **Multi-fenêtrage par croissance** – Dans [Lotti 94], les auteurs supposent que les ruptures de profondeur coïncident avec les contours de l'image. Une première étape consiste à détecter les contours de l'image à l'aide du filtre de Canny-Deriche et un seuillage pas hystérésis. Quatre fenêtres carrées de dimensions 1×1 sont ensuite centrées sur le pixel considéré. Chaque fenêtre est associée à l'une des quatre directions nord-ouest, nord-est, sud-est, sud-ouest et est agrandie progressivement sans qu'elle ne chevauche les contours présents dans la zone étudiée. Sur l'exemple de la figure 2.14, la croissance est réalisée à partir du point marqué en noir et où les lignes épaisses représentent des

FIGURE 2.13: *Multi-fenêtrage fixe.*

contours. Le score de corrélation final est calculé en prenant l'union des quatre fenêtres claires.

FIGURE 2.14: *Multi-fenêtrage par croissance.*

- **Multi-fenêtrage adjacent** – Dans [Hirschmüller 02], une fenêtre de corrélation est centrée sur le pixel considéré puis d'autres fenêtres sont placées autour et la chevauchent. Chaque fenêtre donne un score de corrélation et le score final est composé de la somme des scores les plus élevés. Sur la figure 2.15 sont présentés 2 exemples de configurations possibles sont montrés. En (a), 5 fenêtres de corrélation sont utilisées. Le score de corrélation final est obtenu en prenant la somme des scores issus des 2 fenêtres de corrélation qui donnent les scores les plus élevés (il s'agit des scores des fenêtres jaune et rouge sur cet exemple). En (b), 9 fenêtres de corrélation sont utilisées. Le score de corrélation final est la somme des scores donnés par les 4 fenêtres de corrélation qui donnent les scores les plus élevés (il s'agit des scores obtenus par 4 fenêtres parmi l'orange, la jaune, la grise, la bleue, la violette et la noire).

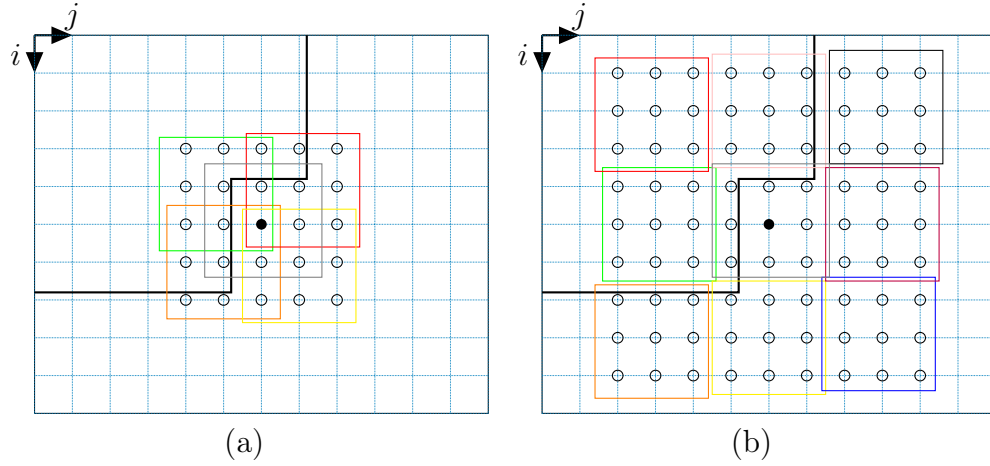


FIGURE 2.15: *Multi-fenêtrage adjacent.*

Déformation indépendante des fenêtres de corrélation

Les méthodes présentées ci-dessus décrivent, pour chaque couple de correspondants, des fenêtres de corrélation de forme identique dans l'image de référence et dans l'autre image sans prendre en compte d'éventuelles déformations sur la partie de l'objet considéré entre les images gauche et droite. Dans ce cas, il est souhaitable de modifier la forme de la fenêtre de corrélation dans l'autre image afin qu'elle prenne en compte les correspondants des pixels inclus dans la fenêtre de corrélation de l'image de référence. Sur la figure 2.16, la fenêtre de corrélation à gauche en noir est transformée par une transformation $\tau^{g \rightarrow d}$ à droite. On marque d'un point bleu le pixel considéré et d'un point violet un de ses voisins dans la fenêtre de corrélation. Pour le pixel considéré, la transformation est donnée par la fonction de disparité d et on constate que le voisin violet ne subit pas la même translation mais subit un déplacement δ . On constate alors que la fenêtre déformée recouvre la même zone de l'objet grisé qu'à gauche. En revanche, si on avait utilisé une fenêtre carrée, en gris, la même partie de l'objet ne serait pas prise en compte, à droite, dans le calcul du score de corrélation.

Soient $\mathbf{p} = [i \ j]^\top$, le vecteur des coordonnées du point considéré $\mathbf{p}_{i,j}^g$ et $(k, l) \in [-n/2; n/2]^2$. Le pixel $\mathbf{p}_{i+k, j+l}^g$ dont les coordonnées sont notées $\mathbf{q} = [i+k \ j+l]^\top$ est un voisin inclus dans la fenêtre de corrélation centrée sur $\mathbf{p}_{i,j}^g$.

On note $\tau^{g \rightarrow d}$ la transformation entre les deux fenêtres de corrélation :

$$\begin{aligned} \tau^{g \rightarrow d} : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ \mathbf{q} &\mapsto \mathbf{q} + \delta(\mathbf{q}) \end{aligned} \tag{2.30}$$

où $\delta(\mathbf{q})$ est une fonction qui exprime le déplacement au point de coordonnées \mathbf{q} :

$$\begin{aligned}\delta : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ \mathbf{q} &\mapsto \delta(\mathbf{q})\end{aligned}\tag{2.31}$$

pour laquelle on peut écrire le développement limité de Taylor au voisinage du point de coordonnées \mathbf{p} :

$$\delta(\mathbf{q}) = \delta(\mathbf{p}) + \frac{\partial \delta(\mathbf{p})}{\partial i} k + \frac{\partial \delta(\mathbf{p})}{\partial j} l + \frac{1}{2} \frac{\partial^2 \delta(\mathbf{p})}{\partial i^2} k^2 + \frac{\partial^2 \delta(\mathbf{p})}{\partial i \partial j} kl + \frac{1}{2} \frac{\partial^2 \delta(\mathbf{p})}{\partial j^2} l^2 + \dots\tag{2.32}$$

Par substitution dans l'équation (2.30), on obtient à l'ordre 0 :

$$\tau^{g \rightarrow d}(\mathbf{q}) = \mathbf{q} + \delta(\mathbf{p}).\tag{2.33}$$

Dans ce cas, il n'y a pas de transformation de la fenêtre de corrélation mais uniquement une translation donnée par le vecteur de disparité du pixel considéré. En revanche, à l'ordre 1, on obtient :

$$\begin{aligned}\tau^{g \rightarrow d}(\mathbf{q}) &= \mathbf{q} + \delta(\mathbf{p}) + \frac{\partial \delta(\mathbf{p})}{\partial i} k + \frac{\partial \delta(\mathbf{p})}{\partial j} l \\ &= \mathbf{q} + \delta(\mathbf{p}) + \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} k \\ l \end{bmatrix}\end{aligned}\tag{2.34}$$

où \mathbf{A} est la matrice des paramètres de la déformation (il s'agit d'une transformation affine) et $\delta(\mathbf{p})$ contient le vecteur des paramètres de la translation. Par déduction, $\delta(\mathbf{p}) = d(\mathbf{p})$ est la fonction de disparité.

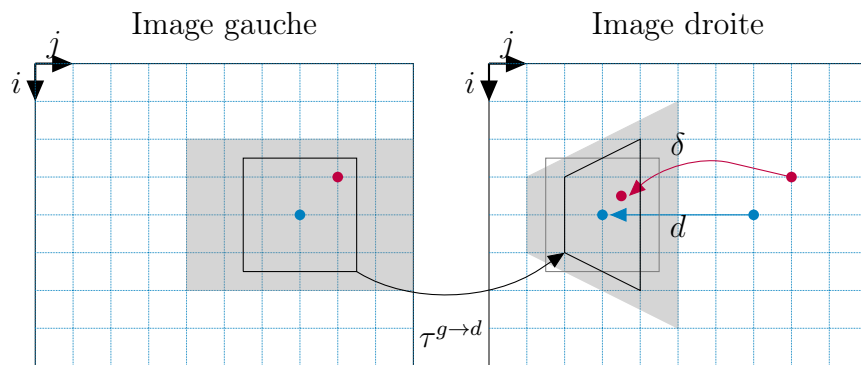


FIGURE 2.16: *Fenêtre adaptative.*

Estimation des paramètres de la transformation affine

La fenêtre s'adapte par une transformation locale estimée entre les images gauche et droite. Cette estimation peut être réalisée :

- **Par optimisation** – Cette approche est utilisée dans [Devernay 94, Garcia 01] pour la mise en correspondance par corrélation fine où les paramètres de la transformation locale de la fenêtre (déformation et translation) sont intégrés à la mesure de corrélation. La mise en correspondance par corrélation fine est également utilisée dans [Harvent 10]. Une idée similaire est utilisée dans [Gruen 85]. Ces paramètres doivent alors maximiser la mesure de corrélation adaptée choisie. Ils sont estimés par une technique d'optimisation, voir § 2.3.3, qui dépend de la mesure de corrélation choisie et qui nécessite une estimation initiale des paramètres. Les paramètres de la translation sont initialement estimés à partir d'une carte de disparité préalablement calculée avec une méthode classique. Les autres paramètres sont initialement considérés nuls. Les vecteurs de disparité résultent donc de l'optimisation de tous ces paramètres.
- **À partir de la distribution des valeurs des pixels dans les voisinages des correspondants** – Dans [Kannala 07], les auteurs estiment au voisinage de chacun des correspondants les paramètres d'une ellipse en fonction de la distribution locale des valeurs des pixels. La transformation affine locale correspond alors à la transformation entre les deux ellipses de deux correspondants. Cette dernière méthode est utilisée pour une mise en correspondance par propagation, comme nous allons le voir § 2.2.4.

Pondération des pixels de la fenêtre de corrélation

Un poids est attribué à chaque pixel de la fenêtre de corrélation centrée sur le pixel considéré. Ce poids donne l'influence de chaque voisin dans le calcul du score de corrélation.

- **Masquage des occultations** – Dans [Kostková 02], une première mise en correspondance avec des fenêtres de corrélation centrées sur les pixels considérés est effectuée. Une seconde mise en correspondance est ensuite réalisée où une nouvelle fenêtre de corrélation est construite pour chaque pixel en fonction des premières disparités estimées de telle sorte que seuls les pixels du voisinage ayant la même disparité estimée que le pixel étudié soient pris en compte dans le calcul du score de corrélation. Il s'agit en fait d'une pondération binaire.
- **Pondération photométrique et spatiale** – Dans [Yoon 05], le poids attribué à chaque pixel de la fenêtre de corrélation est proportionnel à la probabilité que le voisin considéré ait la même disparité que le pixel central. Cette probabilité est estimée en prenant en compte la différence de couleur entre les deux pixels ainsi que leur distance. Dans [Gong 07], une version simplifiée de cette dernière technique est évaluée et comparée à d'autres méthodes d'agrégation dont un fenêtrage carré et différents multi-fenêtrages. La méthode par pondération est celle qui donne le moins d'appariements erronés.

REMARQUE : on peut considérer que les techniques par pondération s'apparentent aux méthodes s'appuyant sur des outils de statistiques robustes. Le lecteur intéressé peut trouver plus de détails dans [Chambon 11].

2.2.4 Mise en correspondance par propagation de germes

Nous avons vu § 2.2.2, que la zone de recherche d'un correspondant peut être réduite à une seule droite en utilisant la géométrie épipolaire. Cette zone de recherche peut encore être réduite en faisant l'hypothèse que, presque partout, les pixels voisins ont des disparités proches. C'est sur cette hypothèse que reposent les méthodes par propagation. Il s'agit d'algorithmes itératifs qui s'appuient sur un ensemble d'appariements fiables appelés des *germes*.

Par la suite, nous utiliserons les termes suivants :

- **Germe** – Un germe est un appariement fiable (couple de pixels). L'ensemble des germes d'un couple d'images, à l'instant t , est noté \mathcal{G}_t (l'ensemble initial est noté \mathcal{G}_0).
- **Pixel gauche d'un germe** – Dans un germe, il s'agit du pixel situé sur l'image de gauche (de référence dans notre cas). Si le pixel $\mathbf{p}_{i,j}^g$ est un pixel de référence d'un germe, alors on le note $\mathbf{g}_{i,j}^g$. On note $\mathcal{R}_{\mathcal{G}_t}^g$, l'ensemble des pixels gauche des germes de \mathcal{G}_t .
- **Pixel droit d'un germe** – Il s'agit du pixel correspondant dans l'image droite du pixel gauche du germe. Si le pixel $\mathbf{p}_{i',j'}^d$ est le correspondant d'un pixel de référence d'un germe, alors on le note $\mathbf{g}_{i',j'}^d$. On note $\mathcal{R}_{\mathcal{G}_t}^d$, l'ensemble des pixels droits des germes de \mathcal{G}_t .
- **Fonction de correspondance des germes** – On définit la fonction de correspondance des germes :

$$\begin{aligned} g^{g \rightarrow d} : \mathcal{R}_{\mathcal{G}_t}^g &\rightarrow \mathcal{R}_{\mathcal{G}_t}^d \\ \mathbf{g}_{i,j}^g &\mapsto \mathbf{g}_{i',j'}^d. \end{aligned} \tag{2.35}$$

- **Fenêtre de voisinage** – Il s'agit de la fenêtre centrée sur le pixel gauche du germe considéré qui inclut les voisins pour lesquels on va chercher un correspondant au voisinage du pixel droit du germe. Pour simplifier, on prend un voisinage carré de dimensions $n_v \times n_v$, voir figure 5.8.
- **Zone de recherche** – Dans le cas général, il s'agit de la zone de recherche des candidats au voisinage du pixel droit du germe considéré. Pour simplifier, on prend une zone de recherche carrée de dimensions n_z^2 . Dans le cas où les images ont été rectifiées, la zone de recherche peut encore être réduite à la zone grisée dans la figure 5.8, seulement constituée de $n_z/2 - 1$ pixels.

Sélection de germes

Nous distinguons deux familles de méthodes de sélection de germes par :

- **appariements de points d'intérêt** : il s'agit d'une mise en correspondance de points d'intérêt ;
- **appariements d'intérêt** : il s'agit de ne conserver que des appariements qui vérifient un ensemble de contraintes.

Ces méthodes sont présentées plus en détails dans le chapitre suivant.

Approches de la propagation

Le principe de la propagation peut s'énoncer ainsi. Pour chaque pixel $\mathbf{p}_{i+k,j+l}^g$, $(k, l) \in [-n_v/2; n_v/2]^2$ voisin du pixel de gauche $\mathbf{g}_{i,j}^g$ du germe considéré, les candidats testés sont ceux de la zone de recherche positionnée autour du voisin du pixel droit de ce germe : $\mathbf{p}_{i'+k+q,j'+l+r}^d$, $(q, r) \in [-n_z/2; n_z/2]^2$, voir figure 5.8. Les nouveaux appariements établis servent ensuite de nouveaux germes et le processus est itéré tant que de nouvelles disparités sont trouvées.

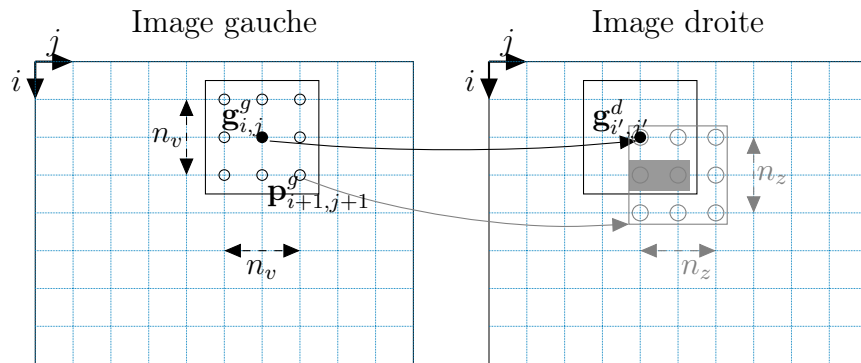


FIGURE 2.17: *Principe de la propagation.* – L'appariement $(\mathbf{g}_{i,j}^g, \mathbf{g}_{i',j'}^d)$ est un germe. En noir, apparaissent les fenêtres de voisinage. Pour les pixels voisins de $\mathbf{g}_{i,j}^g$ dans l'image de gauche (points blancs entourés), la recherche de leurs correspondants s'effectue au voisinage des voisins du correspondant du germe considéré. Plus précisément, si on prend par exemple le pixel $\mathbf{p}_{i+1,j+1}^g$, sa zone de recherche (fenêtre en gris) est constituée des pixels $\mathbf{p}_{i'+1+q,j'+1+r}^d$ avec $(q, r) \in [-n_z/2; nm_z/2] \times [-n_z/2; n_z/2]$. Lorsque les images ont subi une rectification épipolaire, la zone de recherche est réduite au sous-ensemble grisé.

Nous distinguons deux approches selon la façon dont les germes sont utilisés :

- **Approche simultanée** – À chaque itération t , tous les germes de \mathcal{G}_t sont considérés simultanément, voir figure 2.18. Cela implique que, lorsque la disparité d'un pixel est calculée, elle peut être induite par plusieurs germes. Dans ce cas, le meilleur appariement, c'est-à-dire celui qui donne le meilleur score de corrélation et qui vérifie l'ensemble des contraintes imposées, est conservé. Cette approche est proposée dans [Chen 99]. Les différentes étapes de cette approche sont présentées dans l'algorithme 2.2.
- **Approche séquentielle** – À chaque itération t , un seul germe de \mathcal{G}_t est considéré, voir figure 2.19, page 54. Cette approche est proposée dans [Otto 89, Lhuillier 02, Ma 05, Kannala 07, Čech 09, Harvent 10]. Ce germe est sélectionné suivant un critère prédéfini et les appariements trouvés ne sont plus remis en cause par la suite. Cette stratégie permet à l'algorithme de propager moins d'erreurs lorsqu'un germe erroné existe, car il n'est pas *a priori* choisi en priorité par le critère de sélection qui sert à traiter les « meilleurs » germes d'abord. Dans [Lhuillier 02, Čech 09], les auteurs sélectionnent les germes à chaque itération selon le score de corrélation le plus élevé. Dans [Otto 89], la plus grande valeur propre de la matrice de covariance obtenue à partir des disparités estimées est utilisée comme critère de sélection des germes. D'après les auteurs, cette valeur est représentative de l'ordre de grandeur de l'erreur commise (leur méthode s'applique à des images de terrains avec peu de discontinuités de profondeur). Dans [Kannala 07], les auteurs reprennent le modèle de [Lhuillier 02] et l'étendent aux couples d'images *wide-baseline*, c'est-à-dire pris depuis des positions relatives éloignées, en prenant en compte les transformations affines locales des valeurs des pixels dues à la perspective (voir § 2.2.3.2). Pour cela, les auteurs calculent les germes par appariements de points d'intérêt en utilisant le détecteur Harris-Affine qui rajoute au détecteur Harris-Laplace (présenté § 3.1.1, page 88) le calcul des paramètres de la « région affine » que l'on peut représenter par une ellipse qui dépend de la distribution des niveaux de gris dans le voisinage. La transformation entre l'ellipse donnée par le pixel gauche du germe considéré et celle de son correspondant, permet d'estimer la transformation affine locale et donc d'adapter la fenêtre de corrélation en conséquence. Les différentes étapes de l'approche séquentielle sont présentées dans l'algorithme 2.3. La transformation de la fenêtre de corrélation est également prise en compte dans [Harvent 10].

Dans l'approche simultanée, à une itération t , l'ordre dans lequel les germes sont sélectionnés n'influe pas sur le résultat. Toutefois, si un appariement erroné est effectué, cette erreur risque de se propager lors des itérations suivantes. L'approche séquentielle permet de limiter les propagations d'erreurs en traitant en priorité les germes *a priori* les plus fiables. Cependant, si le critère de sélection est mal choisi, on risque de considérer fiables de mauvais germes et donc, là aussi, de propager plus d'erreurs.

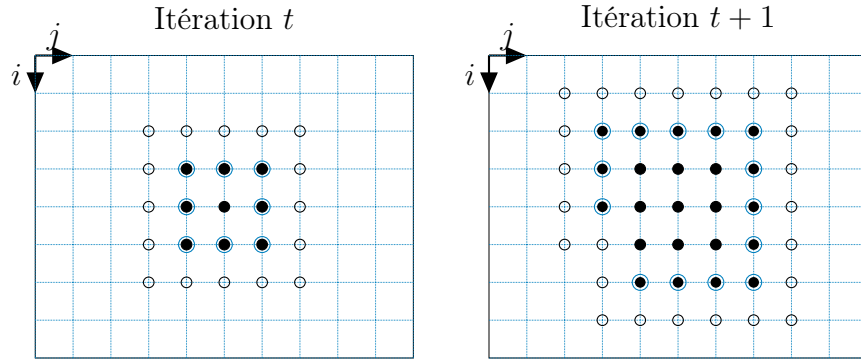


FIGURE 2.18: *Approche simultanée de la propagation.* – Les pixels pour lesquels un correspondant a déjà été trouvé sont marqués par un point noir. Les germes sont les points noirs entourés. Les pixels blancs cerclés sont ceux pour lesquels on recherche un correspondant à l'itération courante. Dans l'approche simultanée, tous les germes sont considérés lors d'une même itération. Sur cet exemple, à l'itération t , on cherche les correspondants des 16 pixels (cerclés blancs), et on en trouve, par exemple, 14 (les germes de l'itération $t+1$).

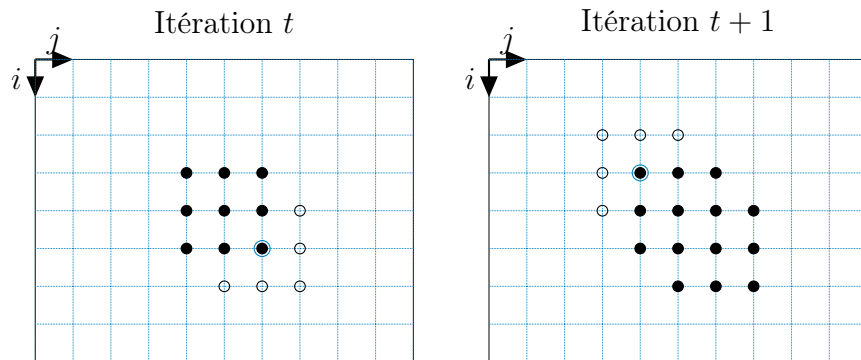


FIGURE 2.19: *Approche séquentielle de la propagation.* – Un seul germe est sélectionné lors d'une même itération (voir la figure 2.18 pour les notations).

```

1:  $\mathcal{G}_t$  : ensemble des germes à l'itération  $t$ 
2:  $\mathcal{M}$  : ensemble des appariements calculés
3:  $t \leftarrow 0$ 
4:  $\mathcal{G}_t \leftarrow$  ensemble initial de germes
5:  $\mathcal{M} \leftarrow \mathcal{G}_t$ 
6: répéter
7:    $\mathcal{G}_{t+1} \leftarrow \emptyset$ 
8:   pour tous les pixels de l'image gauche qui ne sont pas des germes faire
9:     construire, à partir de tous les germes voisins, la zone de recherche, en prenant
       l'union de chaque zone donnée par chacun des germes voisins (voir figure 5.8 pour
       la construction d'une zone de recherche)
10:    rechercher dans la zone de recherche le candidat qui donne le meilleur score de
       corrélation
11:    si le candidat trouvé donne un appariement qui vérifie un ensemble de contraintes
       a posteriori alors
12:      ajouter cet appariement à  $\mathcal{M}$ 
13:      ajouter cet appariement à  $\mathcal{G}_{t+1}$ 
14:    fin si
15:  fin pour
16:   $t \leftarrow t + 1$ 
17: jusqu'à  $\mathcal{G}_{t+1} = \emptyset$ 
18: retourne  $\mathcal{M}$ 

```

ALGORITHME 2.2: *Approche simultanée de la propagation pour la mise en correspondance de pixels.* – À chaque itération, l'ensemble des germes est considéré pour la recherche des correspondants.

```

1:  $\mathcal{G} \leftarrow$  ensemble initial de germes
2:  $\mathcal{M} \leftarrow \mathcal{G}$ 
3: répéter
4:   sélectionner  $\mathbf{g}$  le germe de  $\mathcal{G}$  qui satisfait le mieux le critère de sélection prédéfini
5:    $\mathcal{G} \leftarrow \mathcal{G} \setminus \mathbf{g}$ 
6:   pour chaque pixel non apparié voisin du germe considéré  $\mathbf{g}$  faire
7:     rechercher le candidat dans la zone de recherche définie par le germe et le pixel
       courant qui donne le meilleur score de corrélation
8:     si le candidat trouvé donne un appariement qui vérifie un ensemble de contraintes
       a posteriori alors
9:       ajouter cet appariement à  $\mathcal{M}$ 
10:      ajouter cet appariement à  $\mathcal{G}$ 
11:    fin si
12:  fin pour
13: jusqu'à  $\mathcal{G} = \emptyset$ 
14: retourne  $\mathcal{M}$ 

```

ALGORITHME 2.3: *Approche séquentielle de la propagation pour la mise en correspondance de pixels.* – À chaque itération, un seul germe est considéré. Il est sélectionné suivant un critère prédéfini afin de propager les « meilleurs » germes d'abord.

Contraintes pour la propagation

Un des inconvénients des méthodes par propagation est le risque de propager des erreurs à partir de germes incorrects. Il est possible de limiter cet effet en utilisant des contraintes qui empêchent la propagation lorsque les appariements réalisés sont jugés peu fiables.

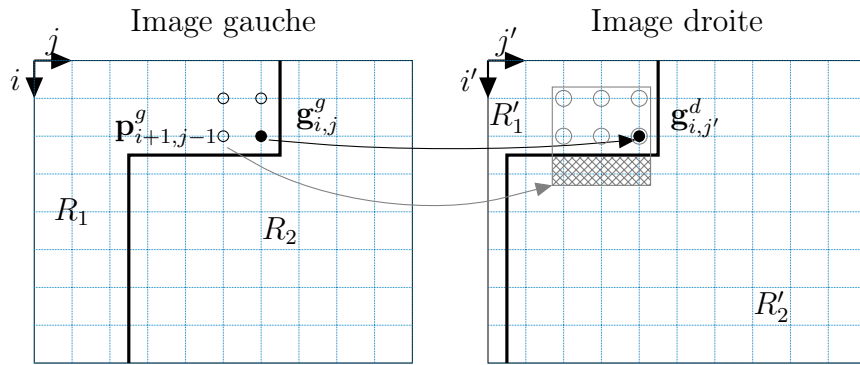
■ **Contrainte d'homogénéité** – La contrainte d'homogénéité a pour but d'empêcher la propagation dans les zones homogènes ou très peu texturées pour lesquelles le manque d'information ne permet pas de trouver des appariements de façon fiable avec une méthode locale [Lhuillier 02]. La propagation d'un germe est stoppée lorsque les quatre pixels voisins de ce germe ont des valeurs proches, c'est-à-dire quand :

$$h(\mathbf{p}_{i,j}^g) = \max \left\{ |I_{\mathcal{I}}^g[i+k \ j+l]^\top) - I^g([i \ j]^\top)| \right\} < t_h \quad (2.36)$$

où $t_h \in \mathbb{R}$ est un seuil et $(k, l) \in \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$.

■ **Contrainte de région** – La contrainte de région suppose que, presque partout, des pixels voisins qui ont des valeurs proches sont les projections de points de la même surface et ont des disparités proches. On suppose alors que les discontinuités de profondeur coïncident avec les contours des régions formées par des pixels de valeurs homogènes. Ainsi, cette contrainte permet d'éviter de propager de mauvaises disparités à l'extérieur

des régions. Sur la figure 2.20, la recherche des correspondants des voisins (cerclés) du pixel gauche $\mathbf{g}_{i,j}^g$ (en noir) appartenant à la région R_1 s'effectue au voisinage du pixel droit $\mathbf{g}_{i,j'}^d$ et la zone de recherche (en gris) se limite aux pixels qui appartiennent à la même région que le pixel droit du germe. Sur cet exemple, les pixels de la région R'_2 sont exclus de la zone de recherche du correspondant du pixel pour le pixel $\mathbf{p}_{i+1,j-1}^g$. Toutefois, l'utilisation de la contrainte de région implique qu'au moins un germe doit être initialement présent dans chaque région pour être capable de trouver les correspondants des pixels de toutes les régions. Cette contrainte est utilisée dans [Ma 05].

FIGURE 2.20: *Contrainte de région.*

2.3 Méthodes globales

Contrairement aux méthodes locales qui prennent en compte uniquement le voisinage de chaque pixel, les méthodes globales cherchent à minimiser un coût évaluant l'erreur globale de mise en correspondance, c'est-à-dire prenant en compte la totalité de l'image. Il s'agit donc de trouver la fonction de correspondance m qui minimise l'erreur globale de mise en correspondance. Cette erreur est quantifiée par une fonction de coût qui intègre un terme d'attache aux données, c'est-à-dire qui vérifie que les valeurs de deux pixels correspondants sont proches, et un terme de lissage qui permet d'assurer la contrainte de la limite du gradient de disparité. Sur la figure 2.21, la carte de disparité est initialisée puis mise à jour tant que le coût global sur l'erreur de mise en correspondance évaluée entre les inconnues (carte de disparité) et les données observées (images gauche et droite)

n'est pas considéré comme minimum. Sur la figure 2.22, on présente l'évolution de la carte de disparité du couples *Cones* durant le processus de la mise en correspondance globale. Ce coût peut être évalué, car, si la carte de disparité cherchée est correcte, alors si on transforme l'image de gauche en fonction de ces disparités, on doit retrouver l'image de droite..

2.3.1 Modélisation du problème à l'aide des champs de Markov

Le problème de mise en correspondance stéréoscopique peut se modéliser à l'aide des champs de Markov, ce qui permet de démontrer que ce problème peut se résoudre par minimisation d'une fonction de coût à l'aide de méthodes d'optimisation spécifiques [Geman 84, Szeliski 90, Li 94, Freeman 00, Veksler 99, Sun 02].

Présentation du problème

Le problème de la mise en correspondance de pixels peut être ramené à un problème d'étiquetage où, dans le cas d'images rectifiées, pour chaque pixel, on souhaite attribuer une valeur de disparité (l'étiquette). Ainsi, si l'image traitée est constituée de hw pixels, et si le nombre de valeurs de disparités possibles est e , alors il existe e^{hw} configurations possibles.

Notations

Par la suite, nous employons les notations suivantes :

- Les valeurs des pixels sont représentées par un ensemble de $n = hw$ variables aléatoires $\mathcal{Y} = \{Y_1, \dots, Y_n\}$.
- Les valeurs des Y_i , $i \in [1; n]$, sont notées y_i .
- On note γ une configuration de \mathcal{Y} .
- Les disparités sont représentées par un ensemble de variables aléatoires $\mathcal{X} = \{X_1, \dots, X_n\}$.
- Les valeurs des X_i sont notées x_i .
- On note χ une configuration de \mathcal{X} .
- On note Ψ l'ensemble des configurations possibles pour \mathcal{X} ($\chi \in \Psi$).

Les valeurs des variables Y_i étant connues, on a donc une configuration donnée γ . L'inconnue du problème est la configuration χ de disparités qui est la plus probable connaissant γ ; on la note $\hat{\chi}$. La solution est donc la configuration $\hat{\chi}$ telle que la probabilité $p(\chi|\gamma)$ soit

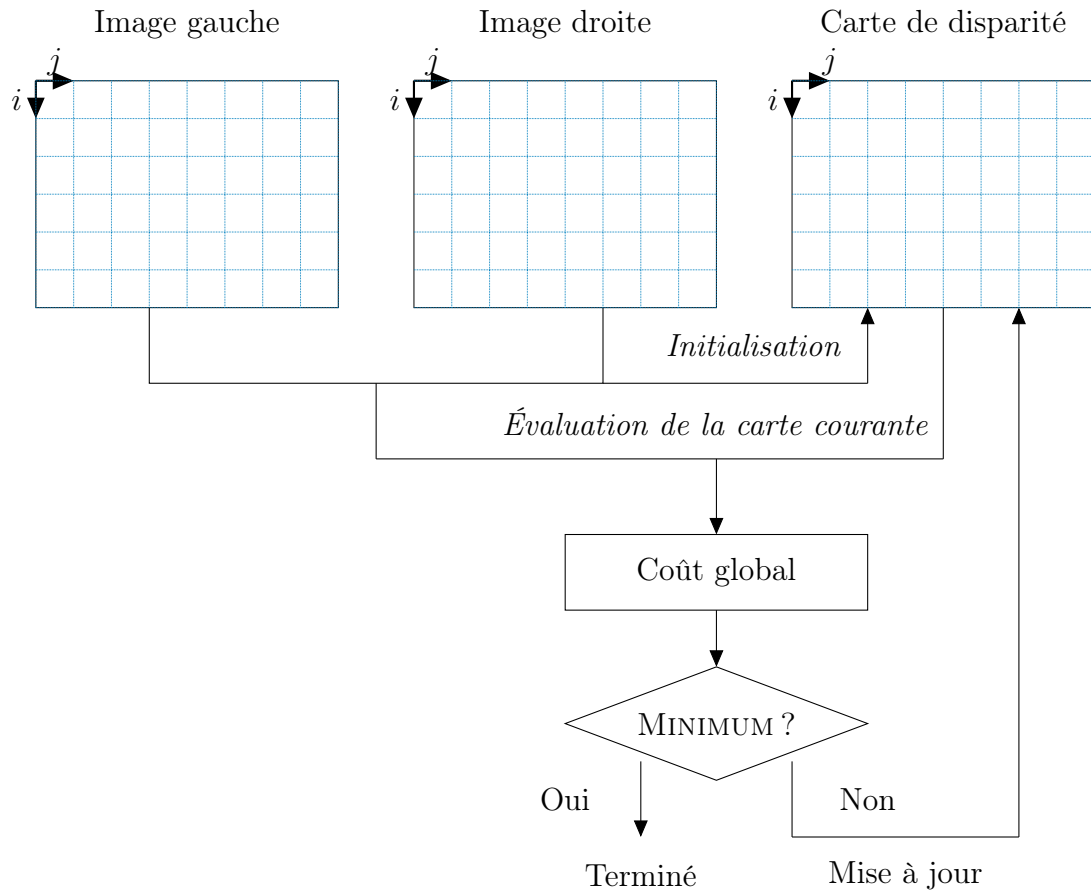


FIGURE 2.21: *Principe des méthodes globales.*

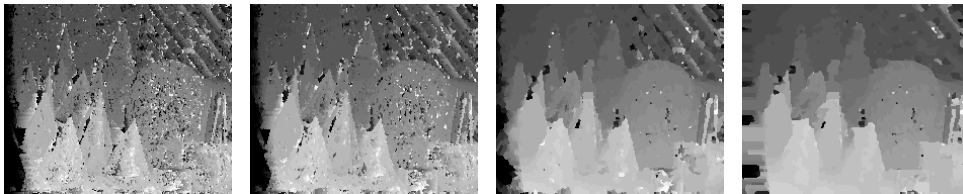


FIGURE 2.22: *Évolution de la carte de disparité du couples Cones durant le processus de la mise en correspondance globale (itérations 0, 2, 20 et 98).*

maximale :

$$\hat{\chi} = \underset{\chi}{\operatorname{argmax}} p(\chi|\gamma). \quad (2.37)$$

C'est ce qu'on appelle le *maximum a posteriori* (MAP). D'après le théorème de Bayes l'équation (2.37) peut se réécrire :

$$\hat{\chi} = \underset{\chi}{\operatorname{argmax}} p(\gamma|\chi)p(\chi). \quad (2.38)$$

Par définition, le graphe qui représente les variables aléatoires de \mathcal{X} constitue un *champ de Markov* car il vérifie les propriétés suivantes :

$$\begin{cases} p(\chi) > 0 & \forall \chi \in \Psi \\ p(x_i|x_j, j \neq i) = p(x_i|x_v) & X_v \in \mathcal{N}_{X_i}^4 \end{cases} \quad (2.39)$$

où $\mathcal{N}_{X_i}^4$ est l'ensemble des variables aléatoires 4-voisines de X_i dans le graphe, voir figure 2.23. Autrement dit, l'état d'une variable aléatoire ne peut dépendre que des états de ses voisins directs.

Pour un champ de Markov, il est alors possible de calculer la probabilité d'une configuration χ grâce au théorème de Hammersley-Clifford qui montre l'équivalence d'un champ de Markov et d'un champ de Gibbs, d'où la probabilité d'une configuration χ :

$$p(\chi) = \frac{1}{z} \exp \left(- \sum_{\{X_p, X_q\} \in \mathcal{N}} V(x_p, x_q) \right) \quad (2.40)$$

où \mathcal{N} contient l'ensemble de tous les couples voisins deux à deux du graphe, et où z est un terme de normalisation où V est une fonction qui quantifie la variation des valeurs x_p et x_q :

$$\begin{aligned} V : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x_p, x_q) &\mapsto V(x_p, x_q). \end{aligned} \quad (2.41)$$

Ayant déterminé $p(\chi)$, pour pouvoir calculer $\hat{\chi}$ d'après l'équation (2.38), il ne reste plus qu'à déterminer $p(\gamma|\chi)$. Il s'agit du terme qui quantifie la probabilité pour un pixel d'obtenir sa valeur connaissant sa disparité. Nous ne pouvons pas le calculer directement. En revanche, si la valeur de disparité est correcte, on doit trouver dans l'autre image un pixel qui a une valeur proche. En calculant la différence des valeurs des deux pixels, on obtient un score inversement proportionnel à ce que l'on souhaite quantifier. On peut donc écrire :

$$p(\gamma|\chi) \propto \exp \left(- \sum_{X_p} D(x_p) \right). \quad (2.42)$$

où \propto signifie que les deux termes sont proportionnels et D est une fonction qui, à partir d'une disparité estimée, quantifie la similarité des deux pixels qu'on suppose correspondants :

$$\begin{aligned} D : \mathbb{R} &\rightarrow \mathbb{R} \\ x_p &\mapsto D(x_p). \end{aligned} \quad (2.43)$$

Par conséquent, d'après les équations (2.38), (2.40) et (2.42), on a :

$$\hat{\chi} = \underset{\chi}{\operatorname{argmax}} \exp \left(- \sum_{X_p \in \mathcal{X}} D_p(x_p) - \sum_{\{X_p, X_q\} \in \mathcal{N}} V(x_p, x_q) \right) \quad (2.44)$$

ce qui revient à chercher la configuration χ qui minimise une *fonction de coût* de la forme :

$$E(\chi) = \sum_{X_p \in \mathcal{X}} D_p(x_p) + \sum_{\{X_p, X_q\} \in \mathcal{N}} V(x_p, x_q). \quad (2.45)$$

Le modèle construit peut être représenté par un graphe, voir figure 2.23.

2.3.2 Fonctions de coût

La fonction de coût qui exprime l'erreur globale de mise en correspondance doit avoir la même forme que la fonction donnée par l'équation (2.45) où chaque variable aléatoire \mathcal{X} correspond à la disparité associée à un pixel. Ainsi on peut considérer qu'une configuration χ peut être représentée par une fonction de disparité d . En outre, les données observées, c'est-à-dire les valeurs des variables aléatoires de \mathcal{Y} , correspondent aux valeurs des pixels des deux images. Par conséquent, nous réécrivons la fonction de coût (2.45) de la manière suivante :

$$\begin{aligned} \underbrace{E(\chi)}_{C(d^g \rightarrow d)} &= \underbrace{\sum_{X_p \in \mathcal{X}} D_p(x_p)}_{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} C_{\text{attache}}(d_{i,j}^g)} + \underbrace{\sum_{\{X_p, X_q\} \in \mathcal{N}} V_{p,q}(x_p, x_q)}_{\lambda \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \sum_{k,l} C_{\text{lissage}}(d_{i,j}^g, d_{i+k,j+l}^g)} \end{aligned} \quad (2.46)$$

avec $(k, l) \in \{(-1, 0), (0, -1), (0, 1), (1, 0)\}$ et où le paramètre $\lambda \in \mathbb{R}$ permet de régler l'influence du terme de lissage par rapport au terme d'attache aux données.

■ **Terme d'attache aux données** – Il est défini par :

$$\begin{aligned} C_{\text{attache}} : \mathbb{R} &\rightarrow \mathbb{R} \\ d_{i,j}^g &\mapsto C_{\text{attache}}(d_{i,j}^g). \end{aligned} \quad (2.47)$$

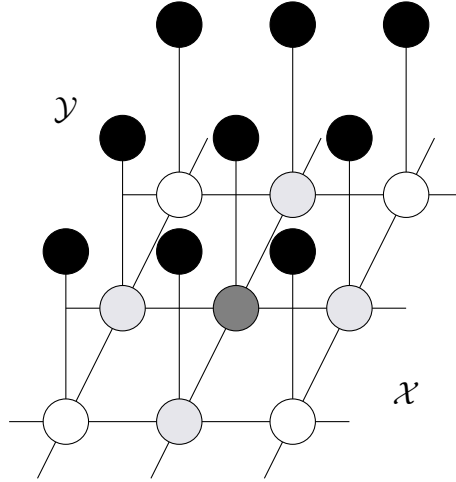


FIGURE 2.23: Modélisation du problème à l'aide des champs de Markov. – Les cercles représentent les variables aléatoires. À chaque cercle noir qui correspond à une variable aléatoire $Y_i \in \mathcal{Y}$ (elle représente la donnée observée), on associe une variable aléatoire $X_i \in \mathcal{X}$ (elles représentent les inconnues). De plus, il existe des liens entre les $X_i \in \mathcal{X}$ voisins qui imposent que les valeurs des variables aléatoires de deux voisins ne soient pas très différentes. Les variables aléatoires \mathcal{X} constituent un champ de Markov car elles sont liées mais chacune ne peut dépendre que de ses voisins directs comme par exemple la variable aléatoire marquée en gris foncé dont l'état ne peut dépendre que de celles marquées en gris clair.

Il évalue l'adéquation entre les disparités données par la fonction de disparité et les données observées. Il correspond généralement à une mesure de corrélation entre les valeurs des pixels de l'image de référence et les valeurs des correspondants donnés par les disparités estimées. Certaines méthodes utilisent la somme de coûts locaux. Par ailleurs, ce terme peut inclure un terme de contraintes. La contrainte la plus fréquemment employée porte sur les occultations. Il s'agit de pénaliser les pixels supposés occultés par une valeur constante prédéfinie λ_{occ} . Dans [Wang 08, Bleyer 10], le terme des occultations est donné par :

$$C_{\text{occ}}(d_{i,j}^g) = \begin{cases} \lambda_{\text{occ}} & \text{si } \mathbf{p}_{i,j}^g \text{ est supposé occulté} \\ 0 & \text{sinon.} \end{cases} \quad (2.48)$$

■ **Terme de lissage** – Ce terme est défini par :

$$C_{\text{lissage}} : \mathbb{R}^2 \rightarrow \mathbb{R} \quad (2.49)$$

$$d_{i,j}^g, d_{i+k,j+l}^g \mapsto C_{\text{lissage}}(d_{i,j}^g, d_{i+k,j+l}^g)$$

Il assure que les disparités de pixels voisins sont proches, ce qui est presque partout le

cas, sauf au niveau des discontinuités de profondeur. Les modèles de fonction de lissage qui sont les plus utilisés dans les problèmes de mise en correspondance sont notamment présentés dans [Veksler 99] :

- **Modèle linéaire** – Une fonction linéaire est utilisée pour mesurer la variation de disparité entre deux pixels voisins :

$$C_{\text{lissage}}(d_{i,j}, d_{i+k,j+l}) = |d_{i,j} - d_{i+k,j+l}| \quad (2.50)$$

avec $(k, l) \in \{(-1, 0), (0, -1), (0, 1), (1, 0)\}$. Ce modèle ne prend pas en compte les discontinuités de profondeur et induit un lissage sur toute la carte de disparité.

- **Modèle linéaire tronqué** – Pour éviter de trop pénaliser le coût global au niveau des discontinuités de profondeur, le modèle linéaire tronqué peut être utilisé :

$$C_{\text{lissage}}(d_{i,j}, d_{i+k,j+l}) = \begin{cases} |d_{i,j} - d_{i+k,j+l}| & \text{si } |d_{i,j} - d_{i+k,j+l}| < \lambda_l \\ \lambda_l & \text{sinon} \end{cases} \quad (2.51)$$

où $\lambda_l \in \mathbb{R}$ est un seuil, voir figure 2.24, page 64.

- **Modèle quadratique tronqué** – Il s'agit d'une alternative au modèle linéaire :

$$C_{\text{lissage}}(d_{i,j}, d_{i+k,j+l}) = \begin{cases} (d_{i,j} - d_{i+k,j+l})^2 & \text{si } (d_{i,j} - d_{i+k,j+l})^2 < \lambda_l \\ \lambda_l & \text{sinon.} \end{cases} \quad (2.52)$$

- **Modèle constant par morceau** – Ce modèle n'impose pas de pénalité pour de faibles variations de disparités entre deux voisins :

$$C_{\text{lissage}}(d_{i,j}, d_{i+k,j+l}) = \begin{cases} 0 & \text{si } |d_{i,j} - d_{i+k,j+l}| < \lambda_l \\ \lambda_l & \text{sinon.} \end{cases} \quad (2.53)$$

2.3.3 Méthodes d'optimisation

Après avoir défini une fonction de coût (2.46), il faut trouver la fonction de disparité d , qui donne le coût le plus faible. En fixant e valeurs de disparité possibles, il existe donc e^{hw} configurations possibles. En pratique, une recherche exhaustive n'est pas envisageable car l'espace des solutions est beaucoup trop grand.

Les méthodes d'optimisation sont des outils qui permettent de trouver le minimum (ou maximum) global d'une fonction, ou bien, une solution proche du minimum recherché. Pour le problème de la minimisation d'une fonction de coût pour la mise en correspondance, les méthodes d'optimisation par *propagation de croyance*, *coupure de graphe*, *recuit simulé*, *algorithme génétique* et *programmation dynamique* sont parmi les plus employées.

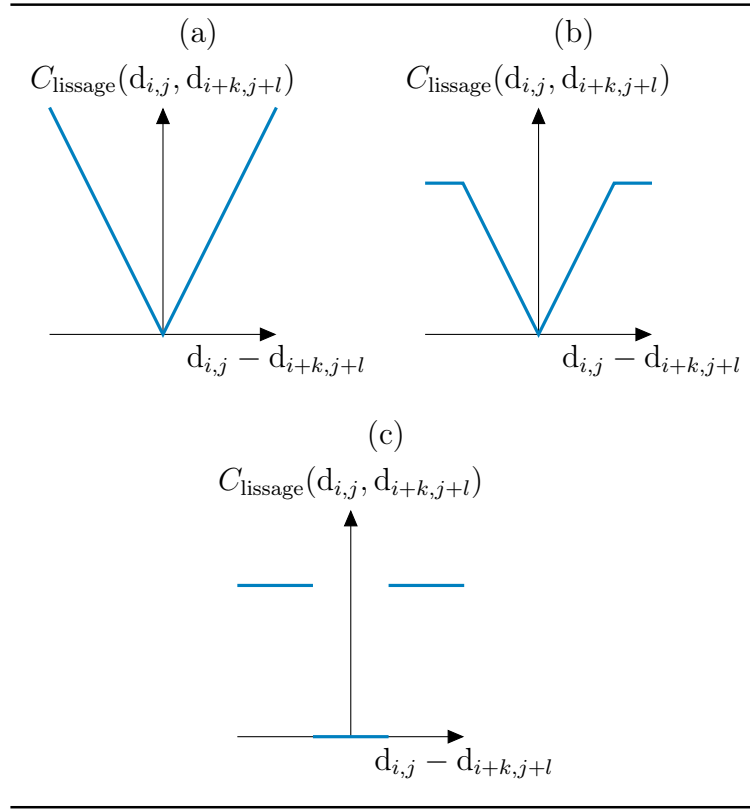


FIGURE 2.24: *Modèles de fonction de lissage.* – (a) *Modèle linéaire*, (b) *modèle linéaire tronqué* et (c) *modèle constant par morceaux*.

- **Propagation de croyance** – La propagation de croyance est un algorithme introduit dans [Pearl 82] permettant de déduire, à partir d'un champ de Markov, la loi marginale en chaque nœud du graphe. Son utilisation pour le problème de la mise en correspondance a été introduite dans [Sun 02]. Elle est également utilisée par [Klaus 06, Yang 09, Xu 08, Yang 08, Sun 05, Zitnick 07, Taguchi 08]. Cette méthode consiste à déterminer pour chaque pixel la densité des disparités et de sélectionner la valeur la plus probable. Pour cela, un algorithme itératif est utilisé. À chaque itération, les pixels voisins s'échangent des messages, sous la forme de vecteurs de dimension égale au nombre de valeurs de disparité possibles. Un message $\mathbf{m}_t^{\mathbf{p}_{i,j}^g \rightarrow \mathbf{p}_{i+k,j+l}^g}$ à

l'itération t entre deux voisins $\mathbf{p}_{i,j}^g$ et $\mathbf{p}_{i+k,j+l}^g$ est donné par :

$$\mathbf{m}_t^{\mathbf{p}_{i,j}^g \rightarrow \mathbf{p}_{i+k,j+l}^g} = \begin{bmatrix} \vdots \\ \left(\min_{d_q \in [d_{\min}; d_{\max}]} \left(C_{\text{lissage}}(d_q, d_p) + C_{\text{attache}}(d_q) + \sum_{k', l'} (\mathbf{m}_{t-1}^{\mathbf{p}_{i+k', j+l'}^g \rightarrow \mathbf{p}_{i,j}^g)_{d_q} \right) \right)_{d_p} \\ \vdots \end{bmatrix} \quad (2.54)$$

où $d_p \in [d_{\min}; d_{\max}]$ est la $p^{\text{ème}}$ valeur de disparité parmi l'ensemble des valeurs possibles et avec $(k, l) \in \{(-1, 0), (0, -1), (0, 1), (1, 0)\}$ et $(k', l') \in \{(-1, 0), (0, -1), (0, 1), (1, 0)\} \setminus \{k, l\}$. Après plusieurs itérations, lorsque les modifications des messages sont très faibles, un vecteur appelé *croyance* est calculé pour chaque nœud :

$$\mathbf{b}^{\mathbf{p}_{i+k,j+l}^g} = \begin{bmatrix} \vdots \\ C_{\text{attache}}(d_p) + \sum_{k, l} \left(\mathbf{m}_t^{\mathbf{p}_{i,j}^g \rightarrow \mathbf{p}_{i+k,j+l}^g} \right)_{d_p} \\ \vdots \end{bmatrix}. \quad (2.55)$$

Enfin, la valeur disparité finale est donnée par :

$$d_{i,j} \leftarrow \underset{d_p}{\operatorname{argmin}} \mathbf{b}^{\mathbf{p}_{i,j}^g}. \quad (2.56)$$

- **Coupure de graphe** – Tout d'abord, le problème que l'on souhaite résoudre est représenté par un graphe orienté. Il existe deux nœuds particuliers : la *source* (aucun arc entrant) et le *puits* (aucun arc sortant). Il s'agit ensuite de trouver dans ce graphe une *coupure minimale*, c'est-à-dire un ensemble d'arcs tel que, sans ces arcs, il n'existe plus dans le graphe de chemin entre le puits et la source et tel que la somme des poids des arcs coupés soit la plus petite possible, ce qui implique un flux maximal entre la source et le puits d'après le théorème de Ford et Fulkerson [Ford 57]. L'algorithme de Ford et Fulkerson permet de déterminer le flux maximal, voir algorithme 2.4. L'algorithme 2.5 introduit dans [Veksler 99] est notamment utilisé pour calculer des disparités, à partir de différents modèles de graphes. Il s'agit d'un algorithme itératif qui s'appuie sur des *move space* qui définissent des changements de configurations. Par exemple, le « $\alpha - \beta$ -relabel move space » consiste, dans la configuration courante, à remplacer la valeur de toutes les variables de valeur α , par la valeur β . Différents modèles de *move space* et de graphes pour des problèmes de vision, dont la mise en correspondance, sont proposés dans [Veksler 99]. Ces modèles diffèrent notamment par le modèle de la fonction utilisée dans le terme de lissage.

Dans [Roy 99], le problème de mise en correspondance est formalisé dans le cas d'une acquisition multi-vues (plus de 2 caméras) et sans utiliser la contrainte épipolaire.

- 1: **répéter**
- 2: Sélectionner au hasard un chemin quelconque de la source au puits (si le poids d'un arc est nul, il n'est pas possible d'emprunter ce chemin).
- 3: Déterminer le flux maximal f_m le long de ce chemin. Il correspond au poids de l'arc le plus faible.
- 4: Soustraire f_m aux poids de tous les arcs du chemin considéré.
- 5: **jusqu'à** ce qu'il n'y ait plus de chemin possible
- 6: Chercher la coupure minimale parmi tous les arcs dont le poids est nul.

ALGORITHME 2.4: *Algorithme de Ford et Fulkerson.* – Il permet de trouver les arcs qui donnent le flux maximal et la coupure minimale d'un graphe.

- 1: $d_0 \leftarrow$ état initial quelconque
- 2: C : fonction de coût
- 3: $\hat{d} \leftarrow d_0$, \hat{d} contient le résultat
- 4: **tant que** vrai **faire**
- 5: Trouver la configuration « voisine » minimale $d_t \leftarrow \underset{d'}{\operatorname{argmin}} C(d')$. Les configurations « voisines » sont données par une fonction de *move space*. Cette configuration minimale est calculée par coupure de graphe.
- 6: **si** $C(d_t) < C(\hat{d})$ **alors**
- 7: $\hat{d} \leftarrow d_t$
- 8: **sinon**
- 9: **retourne** \hat{d}
- 10: **fin si**
- 11: **fin tant que**

ALGORITHME 2.5: *Algorithme de Veksler.*

- **Recuit simulé** – Le recuit simulé est un algorithme itératif et stochastique, c'est-à-dire basé sur des tirages aléatoires, voir algorithme 2.6. Il est notamment utilisé dans [Park 07]. À chaque itération, une nouvelle configuration des disparités est tirée au sort. Si le coût de cette nouvelle configuration est plus faible que le meilleur coût courant, alors elle est acceptée. Dans le cas contraire, elle est acceptée suivant une certaine probabilité qui dépend d'une variable appelée *température*, et cela afin d'éviter que l'algorithme ne converge vers un minima local. En outre, à chaque itération, une fonction fait décroître la température pour que la probabilité d'acceptation devienne de plus en plus faible.

```

1:  $d_0 \leftarrow$  état initial quelconque
2:  $T_0 \leftarrow$  la température initiale
3:  $T \leftarrow T_0$ 
4:  $f_u$  : fonction de disparité courante
5:  $f_d$  : fonction de décroissance de la température
6: répéter
7:    $d_t \leftarrow f_u(d_{t-1})$ 
8:    $T \leftarrow f_d(T)$ 
9:    $\Delta C = C(d_t) - C(d_{t-1})$ , variation de coût
10:  si  $\Delta C < 0$  alors
11:    accepter le changement
12:  sinon
13:    accepter le changement avec une probabilité fonction de la température
14:  fin si
15: jusqu'à convergence
16: retourne  $d_t$ 

```

ALGORITHME 2.6: *Algorithme du recuit simulé pour minimiser une fonction de coût de la fonction de disparité. La fonction f_u met à jour la fonction de disparité courante.*

- **Algorithme génétique** – Cette technique est introduite dans le cas général dans [Holland 75]. Le problème à résoudre est modélisé par une population regroupant un ensemble d'individus. Chaque individu possède une carte génétique, c'est-à-dire un ensemble de propriétés. Les individus les plus adaptés sont ensuite sélectionnés par une fonction de *fitness* et se reproduisent en mélangeant leurs gènes. L'idée est d'obtenir de nouveaux individus avec les meilleures propriétés des deux parents. Des mutations aléatoires peuvent également intervenir. Ainsi, une nouvelle population avec une *fitness* plus élevée est créée. Le processus est réitéré jusqu'à l'émergence d'une population

optimale, c'est-à-dire qui n'évolue plus ou presque plus. Dans le cas de la mise en correspondance, un individu correspond à une fonction de disparité dont les propriétés sont les différentes valeurs de disparité associées à chaque pixel. La sélection des meilleures fonctions de disparité est basée sur la fonction de coût global. Le processus est répété plusieurs fois et il en résulte l'émergence d'une population « optimale ». Enfin, le meilleur individu de la population émergente, c'est-à-dire la meilleure fonction de disparité, est sélectionné comme résultat.

- 1: $\mathcal{D}_0 \leftarrow \{\dots, d_k, \dots\}$, état initial quelconque (génération d'une population initiale de $k = 0, \dots, n$ individus)
- 2: **répéter**
- 3: Sélectionner les m meilleurs individus de la population \mathcal{D} selon la fonction de *fitness*
- 4: Croiser les meilleurs individus entre eux pour engendrer une nouvelle population. Le partitionnement des gènes étant réalisé au hasard
- 5: Effectuer des mutations aléatoires
- 6: **jusqu'à** Émergence d'une population satisfaisant au mieux la *fitness*
- 7: **retourne** le meilleur individu d de la dernière population

ALGORITHME 2.7: *Algorithme génétique.* – Cet algorithme effectue des croisements sur des individus et sélectionne les fils qui satisfont au mieux une fonction de coût (*fitness*).

- **Programmation dynamique** – Cette technique consiste à calculer pour chaque ligne, le chemin qui donne le coût le moins élevé dans une matrice qui répertorie tous les coûts entre tous les appariements possibles de la ligne considérée. Toutefois, cette technique ne prend pas en compte les relations des lignes entre elles. Cette méthode construit pour chaque ligne de l'image de référence une matrice qui contient tous les coûts de correspondance. Il s'agit ensuite de trouver un chemin dans cette matrice qui minimise la fonction de coût.

Le lecteur intéressé pourra trouver des informations complémentaires dans [Scharstein 02, Szeliski 10].

2.4 Méthodes fondées sur les régions

Le principe des méthodes utilisant les régions repose sur l'hypothèse qu'une région de couleur homogène dans l'image suffisamment petite peut être la projection d'une même surface de la scène et que cette surface peut être approchée par un modèle dans l'espace des disparités (par exemple un plan ou une surface *B-Spline*) [Hong 04, Sun 05, Klaus 06, Wang 08, Yang 08, Yang 09, Taguchi 08, Bleyer 10]. L'utilisation de ces régions peut aider la mise en correspondance de régions homogènes et permet également d'éviter des artéfacts au niveau des discontinuités de profondeur. Le modèle plan est relativement simple et donne de bonnes approximations, en particulier avec les plus petites régions, mais peut se révéler imprécis lorsque les régions ne sont pas les projections de facettes planes. Dans ce dernier cas, le modèle *B-Spline* peut donner de meilleures approximations. Cependant, il est plus complexe à configurer (l'ordre et le vecteur nodal doivent être choisis avec soin) et d'importantes oscillations peuvent apparaître dans le résultat [Lin 03, Bleyer 10].

Les méthodes fondées sur les régions comportent plusieurs étapes résumées par la figure 2.25 :

- **Segmentation des images en régions** – L'algorithme *mean-shift* est le plus couramment employé pour réaliser cette opération. Il est détaillé ci-après. Contrairement à la plupart des algorithmes de segmentation, celui-ci ne nécessite pas de spécifier le nombre de régions en entrée. Ce paramètre serait particulièrement difficile à déterminer *a priori* car il dépend notamment de la scène. Par ailleurs, une sur-segmentation est presque toujours calculée. Elle permet d'avoir des petites régions qui sont plus facilement approchées par un modèle plan.
- **Initialisation des disparités** – Une méthode de mise en correspondance classique est généralement employée pour cette étape. En outre, les techniques présentées précédemment (utilisation de contraintes, adaptation de la mesure ou de la fenêtre de corrélation au contexte, voir § 2.2) peuvent être utilisées pour limiter le nombre d'erreurs [Klaus 06, Wang 08, Yang 09].
- **Estimation des paramètres du modèle choisi** – Le modèle plan est relativement simple et donne de bonnes approximations des disparités, en particulier pour les petites régions. Il s'agit du modèle le plus couramment utilisé dans les travaux existants mais le modèle *B-Spline* a également été étudié dans [Lin 03, Bleyer 10]. Les paramètres du modèle choisi sont habituellement estimés à l'aide d'une méthode d'estimation robuste à partir des disparités initiales :
 - ▶ *Moindres carrés pondérés itératifs* – Cette technique est notamment utilisée dans [Hong 04]. À chaque itération, les disparités initiales trop éloignées du modèle estimé sont considérées comme des données aberrantes et sont écartées pour ne pas influencer le résultat.
 - ▶ *RANSAC* – L'algorithme RANSAC (*random sample consensus*) est utilisé dans [Yang 08, Yang 09].

- *Décomposition des paramètres* – Cette technique est employée dans [Wang 08, Klaus 06] pour déterminer les paramètres d'un plan. En écrivant la fonction du plan comme une fonction de i et de j :

$$d_p : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \mapsto p_1 i + p_2 j + p_3 = d_{i,j} \quad (2.57)$$

où p_n (avec $n = 1, \dots, 3$) sont les paramètres recherchés. Les auteurs remarquent que :

$$p_1 = \frac{\partial d_p([i \ j]^\top)}{\partial i} \quad p_2 = \frac{\partial d_p([i \ j]^\top)}{\partial j} \quad p_3 = d_{i,j} - p_1 i - p_2 j. \quad (2.58)$$

Ainsi, p_1 (respectivement p_2), est estimé en prenant toutes les paires possibles de pixels suivant les lignes (respectivement les colonnes) pour estimer les dérivées, chaque paire donnant une estimation pour le paramètre considéré. Un histogramme de ces estimations est alors construit puis lissé par un noyau gaussien. La valeur correspondant au maximum de cet histogramme est la valeur recherchée du paramètre. Enfin, un processus analogue est utilisé pour trouver p_3 à partir de l'ensemble des disparités initiales de la région.

- **Raffinement des résultats** – Une optimisation globale permet de raffiner les résultats. Nous pouvons distinguer les méthodes suivantes :
 - *Affectation d'un plan* – Les différents plans de disparité sont obtenus à partir des modèles estimés et chaque pixel est affecté à l'un de ces plan de telle sorte qu'un coût global sur l'image entière soit optimisé. Cela permet de corriger certaines erreurs notamment dues à une mauvaise segmentation. Cette solution est utilisée dans [Hong 04, Bleyer 10] où la technique d'optimisation employée est la coupure de graphe, [Klaus 06] où la technique d'optimisation employée est la propagation de croyance.
 - *Raffinement des paramètres* – Les paramètres des modèles préalablement estimés peuvent être affinés en minimisant un coût global. Cette solution est notamment utilisée dans [Wang 08, Yang 08].

Segmentation des images en régions

Une partition de l'image est calculée en utilisant un algorithme de segmentation en régions de couleur homogène [Gonzalez 04, Trémeau 04]. Deux approches se distinguent :

- **Approche « contour »** – Les frontières entre les zones de changement brusque des valeurs des pixels, c'est-à-dire les contours de l'image, sont détectés puis utilisés pour délimiter les différentes régions.
- **Approche « région »** – Les pixels voisins qui se ressemblent sont regroupés pour former les régions.

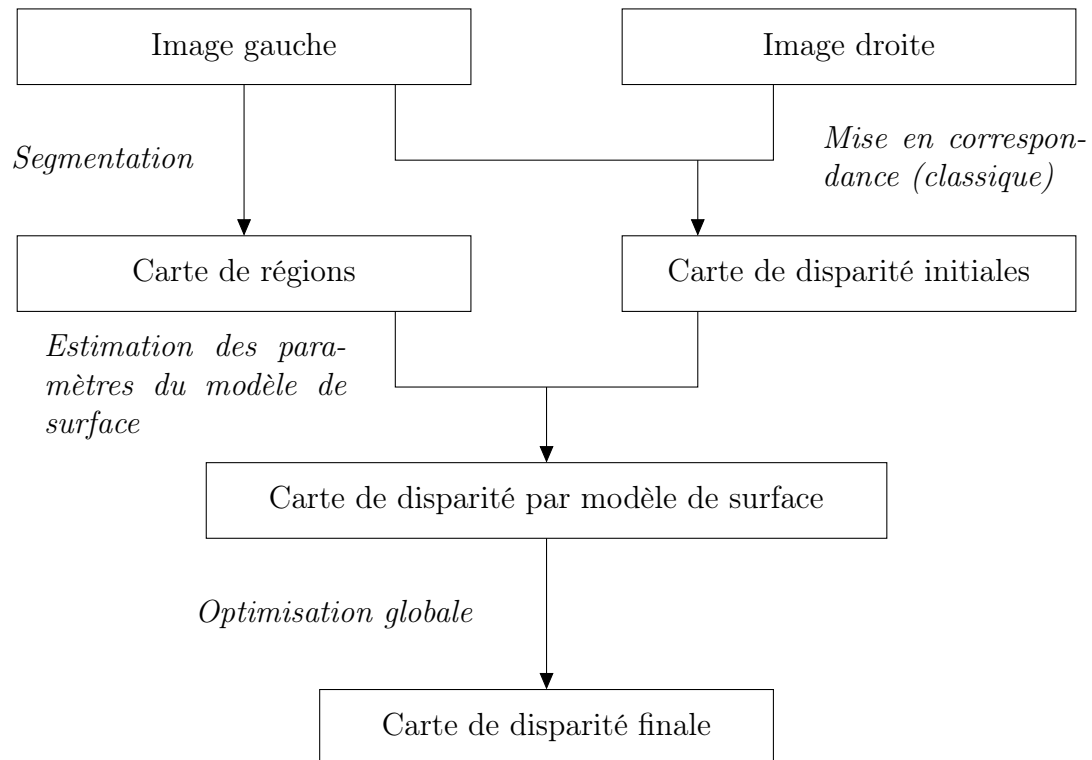


FIGURE 2.25: Étapes des méthodes de mise en correspondance fondées sur les régions.

La méthode *mean-shift* est la plus utilisée dans le contexte de la mise en correspondance fondée sur les régions [Sun 02, Bleyer 05, Klaus 06, Wang 08, Yang 09, Bleyer 10]. Il s'agit d'une approche « région ».

Mean-shift

Nous détaillons ici la méthode de segmentation *mean-shift* proposée dans [Comaniciu 97] qui est largement utilisée pour des problèmes de mise en correspondance. Elle présente l'avantage de ne pas nécessiter la spécification du nombre de régions à trouver. La segmentation est en fait un problème de classification qui consiste à regrouper les pixels voisins ayant des valeurs proches. En termes statistiques, le problème peut être vu comme l'estimation des modes de la fonction de densité de la couleur des pixels.

Densité de probabilité

À partir d'un échantillon, c'est-à-dire d'un sous-ensemble d'observations parmi tous les cas possibles, il est possible d'estimer la loi de probabilité dans le cas discret ou la fonction de densité dans le cas continu. Cette loi ou fonction indique pour chaque valeur possible

d'une variable aléatoire X la probabilité d'obtenir cette valeur.

Dans le cas discret, la probabilité d'obtenir chaque valeur de X est estimée en divisant le nombre d'occurrences de cette valeur par le nombre d'observations.

Dans le cas continu, on considère un intervalle $[a; b]$. Une fonction de densité est définie par :

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto f(x) \end{aligned} \quad (2.59)$$

avec

$$p(a \leq x \leq b) = \int_a^b f(x) dx. \quad (2.60)$$

où $f(x) \geq 0$ et $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Estimation de la densité de probabilité

Pour estimer une fonction de densité à partir d'un échantillon, on distingue deux types de méthodes :

- **les méthodes paramétriques** – lorsqu'un modèle (par exemple gaussien) de la fonction de densité est supposé, le problème revient à estimer les paramètres de ce modèle ;
- **les méthodes non paramétriques** – lorsqu'un modèle n'est pas supposé, des méthodes non paramétriques sont utilisées. C'est le cas de *mean-shift* où la fonction de densité est estimée en utilisant une estimation par noyau.

À partir d'un échantillon de n observations $x_i, i = 1, \dots, n$, la fonction de densité estimée, notée $\hat{f}(x)$, est calculée pour $x \in \mathbb{R}$ en dimension 1 par :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2.61)$$

où h est un paramètre de lissage et K est le noyau utilisé, voir figure 2.26. Nous considérons le noyau d'Epanechnikov (car il permet de simplifier le problème pour donner l'algorithme *mean-shift* détaillé ci-après) et son expression en dimension 1 est la suivante :

$$K_E(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{si } |x| \leq 1 \\ 0 & \text{sinon.} \end{cases} \quad (2.62)$$

Modes d'une densité de probabilité par mean-shift

Les modes de la fonction de densité, c'est-à-dire les valeurs de x où $\hat{f}(x)$ est localement maximale, correspondent aux « centres » des différentes classes cherchées, voir figure 2.26. Le calcul explicite de ces modes en passant par le calcul de la fonction de densité peut être évité en utilisant l'algorithme *mean-shift*. Cet algorithme est itératif. À chaque itération deux étapes sont réalisées :

- **Calcul de la moyenne (*mean*)** – La moyenne est calculée dans une fenêtre de dimensions à définir. Soit x_i les n éléments inclus dans la fenêtre centrée sur l'élément considéré x :

$$\mu(x) = \frac{1}{n} \sum_{i=1}^n x_i - x \quad (2.63)$$

- **Décalage (*shift*)** – Les éléments inclus dans la fenêtre considérée sont translatés sur la moyenne $\mu(x)$.

Après plusieurs itérations, il est alors possible de montrer que chaque élément translaté converge vers le mode le plus proche de la fonction de densité [Comaniciu 97].

Analogie avec le problème de segmentation d'image en régions Le problème de la segmentation d'image en régions de couleur homogène est similaire à celui de la classification des éléphants présentée dans la figure 2.26. Mais l'espace de classification est maintenant à trois ou cinq dimensions au lieu d'être à une dimension : chaque pixel de l'image représente une observation pour laquelle on considère la localisation spatiale, c'est-à-dire les coordonnées du pixel dans l'image, soit deux dimensions, ainsi que le niveau de gris ou la couleur, soit, une ou trois dimensions supplémentaires, voir figure 2.27. Ces dimensions ne sont pas toutes comparables (en particulier les dimensions spatiales et les composantes de la couleur). Par conséquent, une normalisation est souhaitable. Toutefois, en pratique, plutôt que de recourir à une normalisation, les auteurs de [Comaniciu 97] proposent d'utiliser deux paramètres notés h_s et h_r pour définir la fenêtre centrée sur l'élément considéré :

- h_s (*spatial*) correspond à la taille de la fenêtre selon les dimensions spatiales ;
- h_r (*range*) correspond à la taille de la fenêtre selon les dimensions photométriques (niveau de gris ou couleur).

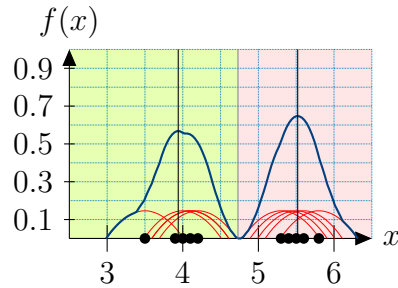
2.5 Conclusion

Nous avons présenté dans ce chapitre différentes techniques de mise en correspondance de pixels pour la stéréovision binoculaire. Nous avons traité plus particulièrement deux approches : la mise en correspondance par propagation de germes et la mise en correspondance fondée sur les régions.

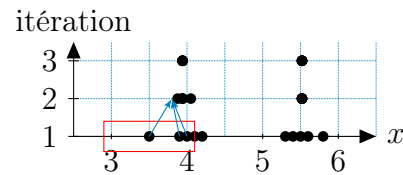
La première est une alternative à la mise en correspondance locale « classique ». Elle s'appuie sur un ensemble d'appariements fiables, les germes, qui permet de réduire certaines ambiguïtés. Mais ces germes doivent être sélectionnés avec soin pour éviter de propager des erreurs. Bien que de nombreux travaux traitent du problème de la mise en correspondance fiable (en particulier pour les points d'intérêt), peu d'entre eux s'intéressent à ce problème pour la sélection de germes dans le cadre de la mise en correspondance pour

ÉLÉPHANT	x_i	$\hat{f}(x_i)$	CLASSE
Bubble	4.0	0.55	C_1
Géant	4.2	0.51	C_1
Bobar	3.9	0.57	C_1
Colonel Défense	4.1	0.54	C_1
Tony	3.5	0.20	C_1
Norton	5.4	0.61	C_2
Palourde	5.8	0.42	C_2
Tantot	5.5	0.65	C_2
Yvoir	5.3	0.52	C_2
Muffin	5.6	0.62	C_2

(a) – échantillon.



(b) – fonction de densité.



(c) – mean-shift.

FIGURE 2.26: Exemple de segmentation par mean-shift. – Le tableau (a) présente les poids x_i (en tonnes) mesurés pour dix éléphants ainsi que les valeurs de la fonction de densité $\hat{f}(x_i)$ correspondante. Ces valeurs sont estimées par la méthode du noyau avec le noyau d'Epanechnikov ((b) en bleu, la fonction de densité; en rouge les noyaux; les x_i sont marqués d'un point noir sur l'axe des abscisses). Les modes en $x_1 = 3.94$ et en $x_2 = 5.52$ sont marqués d'une barre verticale. Ainsi, les éléphants dont le poids est inférieur à $3.94 + (5.52 - 3.94)/2 = 4.73$ tonnes (en vert sur la figure (b)) peuvent être regroupés dans une classe C_1 , ceux dont le poids est supérieur à cette valeur, peuvent être regroupés dans une classe C_2 . La classe de chaque éléphant est reportée dans le tableau (a). La figure (c) montre qu'on peut retrouver ces résultats en utilisant l'algorithme mean-shift par trois itérations. Prenons par exemple le premier point noir sur la gauche. La fenêtre utilisée est marqué en rouge. La moyenne des éléments inclus dans la fenêtre est calculée et sont déplacés sur cette moyenne. Le processus est itéré plusieurs fois jusqu'à convergence vers un point qui correspond au mode le plus proche de la fonction de densité. On considère alors que les points qui ont convergé vers un même mode appartiennent à la même classe (ou région).

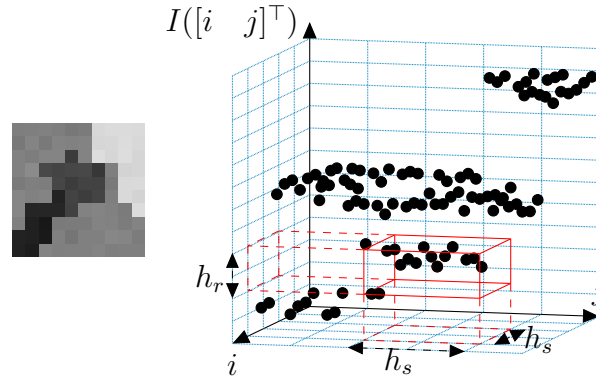


FIGURE 2.27: *Espace de classification 3D.* – Chaque point noir représente un pixel caractérisé par ses coordonnées dans l'image (i et j) ainsi que par son niveau de gris $I([i \ j]^T)$. La couleur est généralement utilisée mais nous utilisons les niveaux de gris pour simplifier cette illustration. La fenêtre mean-shift de dimension (h_s, h_s, h_r) est représentée en rouge.

la stéréovision binoculaire. C'est la raison pour laquelle nous en proposons une étude détaillée dans les deux chapitres suivants.

La seconde approche permet d'obtenir les meilleurs résultats selon le protocole d'évaluation de Middlebury². Nous proposons dans le chapitre 6, une technique de mise en correspondance qui s'inspire et tire profit des avantages de ces deux approches.

2. <http://vision.middlebury.edu/stereo>

3 Sélection des germes initiaux pour la mise en correspondance par propagation

LA mise en correspondance par propagation est une alternative intéressante aux méthodes locales classiques. Cependant, de nombreux aspects de ce type de méthodes ont été peu étudiés, en particulier en ce qui concerne la sélection des germes initiaux. Le fait de réduire la zone de recherche des correspondants au voisinage de germes permet non seulement de réduire les calculs mais également le risque de sélectionner un mauvais correspondant, tout en diminuant les ambiguïtés. Néanmoins, un soin particulier doit être apporté à la sélection de ces germes.

Nous distinguons deux familles de méthodes de sélection des germes :

- **Appariement de points d'intérêt** – Les points d'intérêt sont des pixels particuliers de l'image ayant des propriétés permettant un appariement plus fiable, c'est-à-dire pour lequel le risque de commettre des erreurs de mise en correspondance est limité. Il existe différents détecteurs de points d'intérêt. Un « bon » détecteur est répétable [Schmid 00], c'est-à-dire qu'il doit être capable de détecter des pixels qui sont les projections des mêmes éléments de la scène dans différentes images. Ainsi, lors de la mise en correspondance d'un point d'intérêt, les candidats sont réduits à un petit sous-ensemble de pixels. Cela présente moins d'ambiguïtés dans le choix du correspondant que s'il avait fallu, par exemple, tester tous les pixels de la droite épipolaire correspondante. Cette approche est notamment utilisée dans [Otto 89, Lhuillier 02, Kannala 07].
- **Appariements d'intérêt** – Il n'y a pas de critère de sélection préalable de pixels.

Une méthode locale classique de mise en correspondance est appliquée à tous les pixels de l'image (ou à un sous-ensemble de pixels). La zone de recherche des candidats est alors plus grande qu'en se limitant à des points d'intérêt. Toutefois, un ensemble de contraintes fortes sont utilisées pour éliminer les appariements qui paraissent moins fiables. Les appariements restants, *a priori* les plus sûrs, constituent les germes. Afin de contrôler la répartition des germes sur toute l'image, une technique de *bucketing*, qui consiste à diviser l'image en blocs et à y chercher des germes, est utilisée dans [Chen 99].

La propriété de répétabilité des détecteurs de points d'intérêt assure que des pixels qui sont les projections d'une même entité de la scène sont détectés sur plusieurs images de cette même scène. Ainsi, la détection de points d'intérêt permet de limiter la recherche à un ensemble de quelques candidats parmi lesquels se trouvent très probablement les correspondants. Cette approche nous a donc paru intéressante à explorer pour la sélection des germes pour la mise en correspondance par propagation. Nous en proposons dans ce chapitre une étude approfondie dans le cadre de la sélection de germes pour la mise en correspondance par propagation. Deux étapes sont distinguées :

- **Détection des points d'intérêt** – La première étape permet de détecter les pixels de l'image qui sont *a priori* plus susceptibles que les autres d'être appariés correctement.
- **Appariement des points d'intérêt** – La seconde étape permet de mettre en correspondance les points d'intérêt. Un descripteur permet de caractériser un point d'intérêt. Il s'agit d'un vecteur permettant de « décrire » le point. Ce vecteur peut être composé, par exemple, des niveaux de gris de son voisinage. D'autres descripteurs sont conçus pour être invariants à des transformations géométriques (rotation, changement d'échelle, changement de perspective). Une mesure de distance permet ensuite de comparer les descripteurs. Il existe de nombreux descripteurs dont une étude comparative peut être trouvée dans [Mikolajczyk 05]. Dans le cas de la mise en correspondance binoculaire de pixels pour la stéréovision où les deux images sont prises à des positions peu éloignées (*small-baseline*), il n'y a pas de fortes transformations géométriques entre les deux images. En revanche, nous sommes confrontés au problème des discontinuités de profondeur. Par conséquent, nous nous concentrons sur les méthodes qui ne sont pas invariantes aux transformations géométriques. Ces méthodes reviennent à réaliser des mesures de corrélation de façon analogue à la mise en correspondance de pixels, notamment avec des mesures robustes.

3.1 Détection de points d'intérêt

Un point d'intérêt est un point particulier de l'image qui possède des caractéristiques intéressantes pour une application donnée. Les points d'intérêt sont, par exemple, utilisés

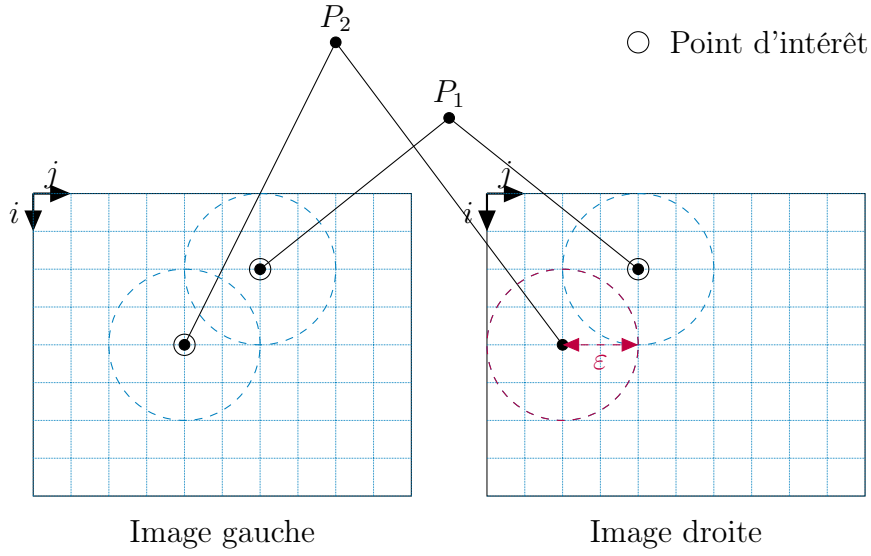


FIGURE 3.1: Répétabilité.

dans le cadre de suivi d'objets dans une séquence d'images [Parisot 09] ou dans la cadre d'indexation [Mikolajczyk 02] et de reconnaissance [Jurie 04]. Dans notre cas, il s'agit d'un point pouvant être mis en correspondance de façon fiable, c'est-à-dire dont les valeurs des voisins présentent des variations caractéristiques. Cela permet de décrire ce voisinage de façon détaillée et de fournir de l'information aidant la mise en correspondance.

Une propriété intéressante des détecteurs de points d'intérêt est la répétabilité. Le critère de répétabilité [Schmid 00] est le rapport entre le nombre de points d'intérêt répétés et le nombre de points d'intérêt détectés. Dans notre cas, un point d'intérêt est dit « répété » si son correspondant est également détecté comme point d'intérêt. Nous ne prendrons volontairement pas en compte les points d'intérêt détectés dans les zones occultées, car par définition, ils ne peuvent pas être répétés. Une marge d'erreur de ϵ pixels est généralement tolérée. Sur la figure 3.1, P_1 et P_2 sont deux points de la scène et les projections de P_1 ont été détectées comme des points d'intérêt sur l'image gauche et sur l'image droite (entourés en noir). Le point de l'image gauche qui est la projection de P_1 est donc répété sur l'image droite, contrairement à la projection de P_2 . Les cercles de rayon ϵ indiquent une marge de tolérance sur l'erreur de détection pour considérer si un point est répété ou pas. Par exemple, si un point d'intérêt avait été présent dans le cercle marqué en violet, alors on aurait considéré la projection de P_2 répétée.

Les détecteurs de points d'intérêt sont utilisés pour leur propriété de robustesse face à différentes transformations photométriques ou géométriques (les points détectés sont répétés malgré la présence de ces transformations entre les deux images). Ces transformations géométriques n'apparaissent généralement pas dans les cas où nous nous plaçons (les deux

caméras ne sont pas dans des positions très éloignées). Les problèmes dominants auxquels nous sommes confrontés sont les occultations et les discontinuités de profondeur.

Dans [Tuytelaars 08], les différents détecteurs de points d'intérêt sont classés suivant le type de point d'intérêt recherché :

- **Coins** – Un coin correspond à l'intersection de plusieurs arêtes d'un objet (jonctions « L », « T » ou « Y »). Un coin détecté ne correspond pas nécessairement à un « coin » d'un objet de la scène. On peut toutefois penser que des points d'intérêt situés sur les coins (c'est-à-dire aux frontières d'objets) risquent d'être proches des ruptures de profondeur ou des occultations.
- **Blobs** – Littéralement un *blob* se traduit par pâté, tache ou grosse goutte [rob 94], c'est-à-dire des termes qui définissent une structure plutôt informe. Une définition formelle est donnée dans [Lindeberg 93] et nous pouvons la résumer de la manière suivante : il s'agit d'une zone de l'image autour d'un point (d'intérêt) où les valeurs des pixels se détachent des valeurs des pixels du voisinage. Par exemple, une zone d'une image qui représente une tache claire située dans une zone sombre constitue un *blob*. En plus du centre du *blob*, un détecteur de ce type doit également caractériser la taille de la zone concernée.

Dans [Schmid 00], les différents détecteurs de points d'intérêt sont classés selon la technique de détection :

- **Détecteurs basés sur les contours** – Ces méthodes détectent les coins au niveau des intersections de contours ou bien au niveau des fortes courbures des contours.
- **Détecteurs basés sur des modèles paramétriques** – Ces méthodes estiment les paramètres d'un modèle de coin pour qu'il s'ajuste précisément sur l'image au niveau du point considéré.
- **Détecteurs basés sur les intensités** – Ces méthodes calculent pour chaque pixel une « valeur d'intérêt », appelée réponse, qui prend en compte les valeurs des pixels du voisinage.

Nous nous sommes intéressés à ces dernières méthodes qui sont les plus utilisées dans les travaux existants. Ces différents détecteurs s'appuient sur les étapes suivantes :

1. **Calcul pour chaque pixel d'une réponse** – Cette réponse dépend des valeurs des pixels au voisinage du pixel considéré. Elle peut utiliser des dérivées, des critères morphologiques ou encore l'entropie de la distribution des valeurs des pixels. Les dimensions du voisinage pris en compte sont liées à une notion d'*échelle*. Lorsque ces dimensions sont fixes, on parle de *détecteur à échelle fixe*. Dans le cas où ces dimensions sont variables, on parle alors de *détecteur multi-échelle*. On note R la fonction qui donne la réponse :

$$R : \mathbb{N}^2 \rightarrow \mathbb{R}^n$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \mapsto R \left(\begin{bmatrix} i \\ j \end{bmatrix} \right). \quad (3.1)$$

Cette fonction peut donner une seule valeur de réponse ($n = 1$) dans le cas des détecteurs à échelle fixe ou bien plusieurs valeurs ($n > 1$) dans le cas des détecteurs multi-échelle (une valeur par échelle). Pour simplifier les notations, on note dans la suite de ce document $\mathbf{p} = [i \ j]^\top$.

2. **Sélection des extrema locaux de la réponse** – Les positions des extrema de la réponse sont localisés. Ces positions constituent des points d'intérêt potentiels.
3. **Post-traitement** – Il s'agit des techniques visant à éliminer les fausses réponses, comme un seuillage sur la valeur de la réponse, la suppression des candidats situés sur des contours ou ne vérifiant pas certaines contraintes géométriques. Il peut également s'agir de techniques permettant un positionnement au sous-pixel des points d'intérêt auquel cas la fonction qui donne la réponse est définie de \mathbb{R}^2 dans \mathbb{R}^n .

Notion d'échelle

L'échelle correspond au niveau de détail observé. Plus le voisinage pris en compte est petit, plus on se focalise sur les détails (voire le bruit) ; inversement, avec un voisinage étendu, on prend en compte les grandes structures au risque de ne pas considérer certains détails. Lorsqu'un détecteur examine plusieurs échelles, la réponse et ses extrema sont calculés dans un « espace échelle ». Il s'agit d'une représentation multi-résolution d'une image telle que plus l'échelle diminue, plus apparaissent de nouveaux détails tout en conservant ceux existants aux échelles supérieures. Il a été montré que seul le lissage gaussien est capable de satisfaire cette contrainte [Babaud 86, Lindeberg 90, Lindeberg 93, Lindeberg 98]. En outre, il vérifie des propriétés de linéarité et d'invariance à la translation. La fonction gaussienne 2D centrée et d'écart-type σ est donnée par :

$$g_\sigma : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\mathbf{p} \mapsto \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}}. \quad (3.2)$$

Sur l'exemple de la figure 3.2, la courbe noire (en haut) représente le signal (interpolé) d'une image 1D pour les pixels de l'axe $i = [0; 50]$. Les courbes colorées (en bas) correspondent à ses dérivées premières dans l'espace échelle. Les dérivées permettent de caractériser les variations du signal initial. Elles sont obtenues par convolution discrète avec un filtre gaussien. Pour chacune de ces deux courbes, l'échelle est donnée par l'écart-type σ de la gaussienne utilisée. Pour la courbe violette on a utilisé $\sigma = 2.25$ et pour la rouge $\sigma = 5.25$. On constate alors, si on s'intéresse aux extrema locaux de la dérivée, que plus l'échelle est grande, moins il y a d'extrema. On ne s'intéresse alors qu'aux plus grandes variations du signal original. On remarque également que les extrema de la dérivée à la plus grande échelle sont aussi des extrema de la dérivée à la plus petite échelle.

On considère, pour chacun des détecteurs décrits ci-après, que l'échelle correspond à la taille du voisinage pris en compte dans le calcul de la réponse. Pour certains détecteurs,

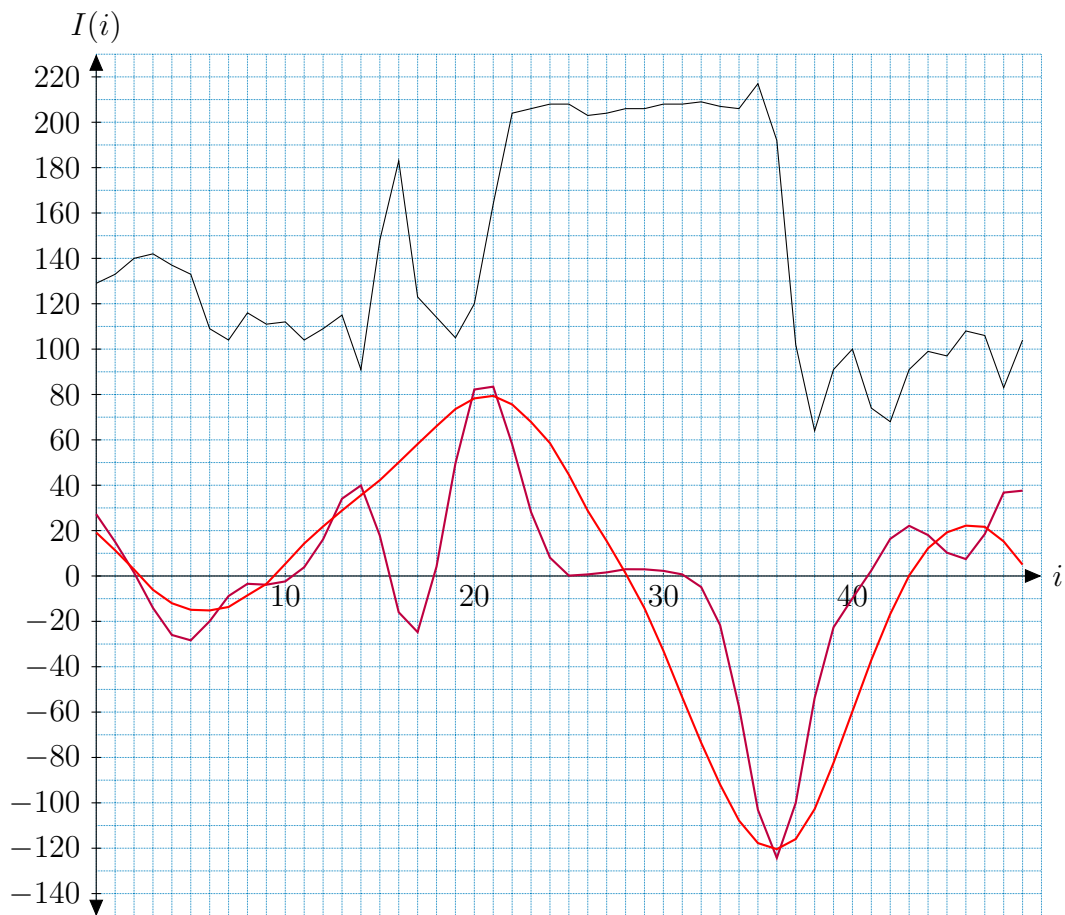


FIGURE 3.2: Exemples d'échelles différentes. – La première courbe représente un signal 1D interpolé et les courbes en dessous correspondent à sa dérivée à deux échelles différentes (courbe violette $\sigma = 2.25$, courbe rouge $\sigma = 5.25$).

cette taille est exprimée sous la forme des dimensions $n \times n$ (en pixels) de la fenêtre carrée centrée sur le point considéré. Pour d'autres détecteurs, la taille est exprimée par le rayon r (en pixels) du voisinage circulaire pris en compte. Enfin, lorsqu'un lissage gaussien est appliqué, l'échelle est donnée par l'écart-type σ . On prendra $n = 4\sigma$ sachant que la surface sous la courbe gaussienne est située à 95% dans l'intervalle $[-2\sigma, 2\sigma]$.

Il existe de nombreux détecteurs. Nous décrivons ici ceux que nous avons le plus rencontrés dans les travaux existants et qui seront utilisés dans la suite pour notre étude comparative. Nous regroupons ces différents détecteurs selon trois familles que nous allons décrire ci-après :

- famille dérivée ;
- famille morphologique ;
- famille entropique.

3.1.1 Famille dérivée

Le calcul de la réponse est fondé sur la variation des valeurs des pixels au voisinage du point considéré.

Réponses de détecteurs à échelle fixe

- **Moravec** – (MO) [Moravec 77] – La réponse prend en compte les variations des valeurs des pixels au voisinage du pixel considéré suivant les directions horizontales, verticales et diagonales (voir figure 3.3) et elle est donnée par :

$$\text{MO}(\mathbf{p}) = \min\{d(\mathbf{p}, [0 \ 1]^\top), d(\mathbf{p}, [1 \ 0]^\top), d(\mathbf{p}, [1 \ 1]^\top), d(\mathbf{p}, [1 \ -1]^\top)\} \quad (3.3)$$

avec

$$d : \mathbb{N}^2, \mathbb{N}^2 \rightarrow \mathbb{R} \\ \mathbf{p}, \mathbf{q} \mapsto \sum_{k,l} (I(\mathbf{p} + [k \ l]^\top) - I(\mathbf{p} + \mathbf{q} + [k \ l]^\top))^2 \quad (3.4)$$

où d est une mesure de dissemblance (SSD) entre un pixel et ses voisins, $n^2 = (2r + 1) \times (2r + 1)$ est la taille du voisinage considéré et $(k, l) \in [-r; r]^2$.

REMARQUE : dans la méthode originale proposée par Moravec, les fenêtres sont de dimensions paires et la réponse est calculée pour la position « subpixelique » située au centre de la fenêtre. En pratique, nous utiliserons des tailles impaires de fenêtres centrées sur le pixel considéré.

■ 3 Sélection des germes initiaux pour la mise en correspondance par propagation

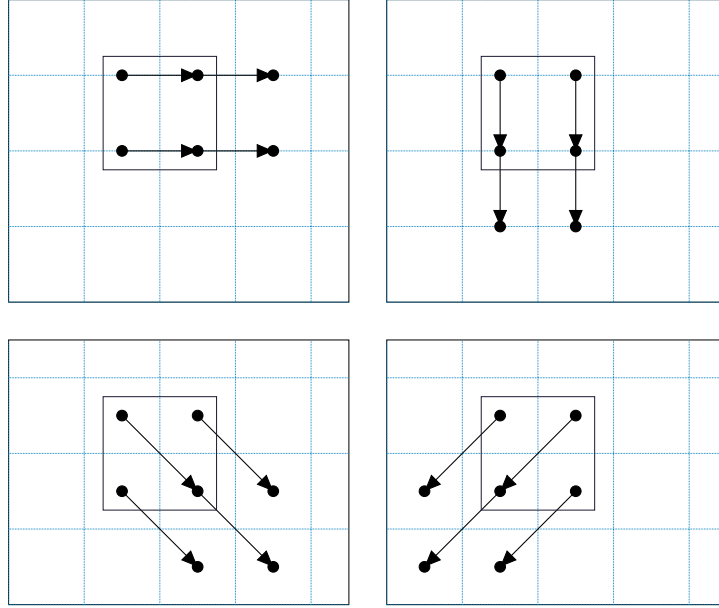


FIGURE 3.3: Détecteur de Moravec. – Pour les quatre directions testées, les flèches indiquent les différentes variations des valeurs des pixels prises en compte dans le calcul de la réponse. Ici une fenêtre de taille 2×2 est utilisée pour la somme des différences au carré. La réponse est donnée par la plus petite des quatre sommes.

- **Harris** – (HA) [Harris 88] – La réponse est calculée pour chaque pixel à partir de la matrice de Harris, parfois appelée matrice des moments du second ordre. Elle généralise une mesure d'auto-corrélation à toutes les directions pour obtenir une réponse isotrope, c'est-à-dire indépendante de la direction. Cette matrice est calculée pour chaque pixel à partir des dérivées premières. Différentes variantes ou améliorations ont été proposées [Harris 88, Noble 88, Shi 94, Gouet 00, Mikolajczyk 05]. Soit $\mathbf{M}_{\mathbf{p}}$ la matrice calculée pour un pixel de coordonnées \mathbf{p} :

$$\begin{aligned} \mathbf{M}_{\mathbf{p}} &= \sigma_D^2 \begin{bmatrix} (g_{\sigma_I} * I_i^2)(\mathbf{p}) & (g_{\sigma_I} * I_i I_j)(\mathbf{p}) \\ (g_{\sigma_I} * I_i I_j)(\mathbf{p}) & (g_{\sigma_I} * I_j^2)(\mathbf{p}) \end{bmatrix} \\ &= \begin{bmatrix} a & c \\ c & b \end{bmatrix} \end{aligned} \quad (3.5)$$

où $*$ est l'opérateur de convolution discrète, g_{σ_I} le noyau gaussien d'écart-type σ_I pour le lissage des valeurs calculées (*integration*) et avec :

$$I_i = \frac{\partial g_{\sigma_D}}{\partial i} * I \quad I_j = \frac{\partial g_{\sigma_D}}{\partial j} * I \quad (3.6)$$

où σ_D est l'écart-type du noyau gaussien utilisé pour le lissage lors du calcul des dérivées (*differentiation*).

Soient $\lambda_1(\mathbf{M}_p)$ et $\lambda_2(\mathbf{M}_p)$ les valeurs propres de la matrice \mathbf{M}_p . D'après [Harris 88] :

- ▶ si les deux valeurs propres sont petites, la région considérée est homogène ;
- ▶ si une des deux valeurs propres est élevée, la zone considérée se situe sur un contour ;
- ▶ si les deux valeurs propres sont élevées, la zone considérée se situe au niveau d'un coin. Dans ce cas, on supposera qu'on est en présence d'un point d'intérêt.

Dans [Shi 94], les auteurs proposent de prendre comme réponse :

$$\text{HA}_1(\mathbf{p}) = \min \{ \lambda_1(\mathbf{M}_p), \lambda_2(\mathbf{M}_p) \}. \quad (3.7)$$

Cependant, le calcul des valeurs propres peut être évité en passant par le calcul de la trace et du déterminant de \mathbf{M}_p :

$$\text{Tr}(\mathbf{M}_p) = \lambda_1(\mathbf{M}_p) + \lambda_2(\mathbf{M}_p) = a + b \quad (3.8)$$

$$\text{Det}(\mathbf{M}_p) = \lambda_1(\mathbf{M}_p)\lambda_2(\mathbf{M}_p) = ab - c^2 \quad (3.9)$$

Une autre réponse fréquemment utilisée est donnée par :

$$\text{HA}_2(\mathbf{p}) = \text{Det}(\mathbf{M}_p) - k\text{Tr}(\mathbf{M}_p)^2 \quad (3.10)$$

où $k = 0.04$. Dans [Noble 88], une alternative pour calculer la réponse à partir de \mathbf{M}_p est proposée :

$$\text{HA}_3(\mathbf{p}) = \frac{\text{Det}(\mathbf{M}_p)}{\text{Tr}(\mathbf{M}_p)}. \quad (3.11)$$

- **Kitchen-Rosenfeld** – (KR) [Kitchen 82] – La réponse est déterminée par la courbure du contour passant par le point considéré pondérée par la norme du gradient en ce point. La réponse est donnée par :

$$\text{KR}(\mathbf{p}) = \left(\frac{I_{ii}I_j^2 - 2I_{ij}I_iI_j + I_{jj}I_i^2}{I_i^2 + I_j^2} \right) (\mathbf{p}) \quad (3.12)$$

où

$$I_i = \frac{\partial I}{\partial i} \quad I_j = \frac{\partial I}{\partial j} \quad I_{ii} = \frac{\partial^2 I}{\partial i^2} \quad I_{jj} = \frac{\partial^2 I}{\partial j^2} \quad I_{ij} = \frac{\partial^2 I}{\partial i \partial j}. \quad (3.13)$$

REMARQUE : les dérivées sont estimées au sens des moindres carrés à partir du développement limité de Taylor des valeurs des pixels jusqu'au second ordre.

- **Beaudet** – (BE) [Beaudet 78] – La réponse utilise les dérivées secondes. Elle est calculée pour chaque pixel à partir du déterminant de la matrice hessienne qui représente la courbure de la surface constituée par les valeurs de l'image au pixel considéré :

$$\text{BE}(\mathbf{p}) = \det(\mathbf{H}_p) = \det \left(\begin{bmatrix} I_{ii}(\mathbf{p}) & I_{ij}(\mathbf{p}) \\ I_{ij}(\mathbf{p}) & I_{jj}(\mathbf{p}) \end{bmatrix} \right). \quad (3.14)$$

- **MIC** (*Minimum Intensity Change*) – (MIC) [Trajkovic 98] – La réponse est donnée par la variation des valeurs des pixels diamétralement opposés par rapport au point étudié. Soit v cette mesure de variation :

$$v : \mathbb{R}^2, \mathbb{R}^2, \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\mathbf{p}, \mathbf{q}, \mathbf{r} \mapsto (I(\mathbf{p} + \mathbf{q}) - I(\mathbf{p}))^2 + (I(\mathbf{p} + \mathbf{r}) - I(\mathbf{p}))^2. \quad (3.15)$$

La variation est d'abord mesurée suivant les directions horizontales et verticales. On pose $v_h = v(\mathbf{p}, [0 \ r]^\top, [0 \ -r]^\top)$ et $v_v = v(\mathbf{p}, [r \ 0]^\top, [-r \ 0]^\top)$ où \mathbf{p} est le vecteur des coordonnées du pixel considéré et r indique l'échelle (en pixels).

Soit $v_1 = \min(v_h, v_v)$. Si cette valeur est en dessous d'un seuil t , la variation est mesurée suivant les autres directions où les valeurs des pixels sont interpolées. Deux modèles de réponse sont proposés en fonction du type de voisinage interpolé considéré, voir figure 3.4 :

- *Voisinage interpolé carré.* – Soit $[k \ l]^\top$ les coordonnées relatives au pixel considéré des pixels situés sur le périmètre du voisinage pris en compte : $r = |k - i| + |l - j|$. Soit la fonction de variation :

$$v' : \mathbb{R}^2, \mathbb{R}^2, \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\mathbf{p}, \mathbf{q}, \mathbf{r} \mapsto (I(\mathbf{p} + \mathbf{r}) - I(\mathbf{p} + \mathbf{q}))(I(\mathbf{p} + \mathbf{q}) - I(\mathbf{p})) + (I(\mathbf{p} - \mathbf{r}) - I(\mathbf{p} - \mathbf{q}))(I(\mathbf{p} - \mathbf{q}) - I(\mathbf{p})). \quad (3.16)$$

On pose $a = \min(v'(\mathbf{p}, [0 \ r]^\top, [r \ 0]^\top), v'(\mathbf{p}, [0 \ -r]^\top, [r \ 0]^\top))$ et $b = v_v - v_h - 2a$. La réponse est ensuite calculée par :

$$\text{MIC}_1(\mathbf{p}) = \begin{cases} v_1 & \text{si } v_1 \geq t \text{ et si } (a \geq 0 \text{ ou si } a + b \leq 0) \\ v_h - \frac{a^2}{v_v} & \text{si } v_1 \geq t \text{ et si } (a < 0 \text{ et si } a + b > 0) \\ 0 & \text{sinon.} \end{cases} \quad (3.17)$$

- *Voisinage interpolé circulaire.* – Soit $[k \ l]^\top$ les coordonnées relatives au pixel considéré des pixels situés sur le périmètre du voisinage pris en compte : $r^2 = (k - i)^2 + (l - j)^2$. Soit la fonction de variation :

$$v'' : \mathbb{R}^2, \mathbb{R}^2, \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\mathbf{p}, \mathbf{q}, \mathbf{r} \mapsto (I(\mathbf{p} + \mathbf{q}) - I(\mathbf{p}))(I(\mathbf{p} + \mathbf{r}) - I(\mathbf{p})) + (I(\mathbf{p} - \mathbf{q}) - I(\mathbf{p}))(I(\mathbf{p} - \mathbf{r}) - I(\mathbf{p})). \quad (3.18)$$

On pose $a = \min(v''(\mathbf{p}, [0 \ r]^\top, [r \ 0]^\top), v''(\mathbf{p}, [0 \ -r]^\top, [r \ 0]^\top))$ et $b = v_h - v_v/2$. La réponse est donnée par :

$$\text{MIC}_2(\mathbf{p}) = \begin{cases} v_1 & \text{si } v_1 \geq t \text{ et si } a \geq 0 \\ v_h - \sqrt{a^2 + b^2} & \text{si } v_1 \geq t \text{ et si } a < 0 \\ 0 & \text{sinon.} \end{cases} \quad (3.19)$$

En pratique, ce détecteur utilise une implémentation multi-résolution, c'est-à-dire que le détecteur est d'abord appliqué à une image de faible résolution, puis seuls les candidats dont le score est supérieur à un seuil t sont considérés lors du calcul de la réponse pour l'image à pleine résolution.

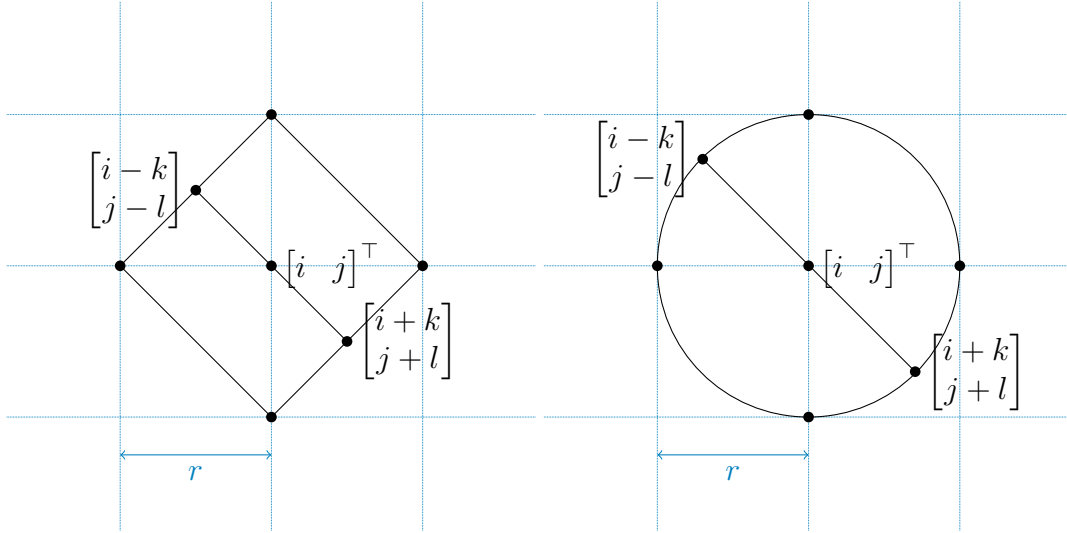


FIGURE 3.4: Détecteur MIC. – Deux modèles d'interpolation sont proposés : le modèle carré (ici à gauche) et le modèle circulaire (à droite).

Réponses de détecteurs multi-échelle

- **SIFT** – (SI) [Lowe 99] – La réponse de ce détecteur repose sur le laplacien des valeurs des pixels à différentes échelles :

$$\text{SI}(\mathbf{p}) = \begin{bmatrix} (\nabla^2 g_{\sigma_0} * I)(\mathbf{p}) \\ \vdots \\ (\nabla^2 g_{\sigma_k} * I)(\mathbf{p}) \\ \vdots \end{bmatrix} \quad (3.20)$$

où σ_0 représente l'échelle initiale et où chaque élément du vecteur $\text{SI}(\mathbf{p})$ correspond à une échelle donnée. On appelle une *octave* un ensemble d'échelles consécutives comprises dans l'intervalle $[\sigma_a; \sigma_b]$ tel que $\sigma_b = 2\sigma_a$. On note o le nombre d'octaves souhaité et

e le nombre d'échelles par octave. En pratique, les auteurs proposent d'approcher ces laplaciens à l'aide de différences de gaussiennes :

$$\text{SI}(\mathbf{p}) \approx \begin{bmatrix} \text{SI}(\mathbf{p})_{\sigma_0} \\ \vdots \\ \text{SI}(\mathbf{p})_{\sigma_k} \\ \vdots \end{bmatrix} \quad (3.21)$$

avec

$$\begin{aligned} \text{SI}(\mathbf{p})_{\sigma_k} &= (g_{\sigma_{k+1}} * I - g_{\sigma_k} * I)(\mathbf{p}) \\ &= (g_{c\sigma_k} * I - g_{\sigma_k} * I)(\mathbf{p}) \\ &= (g_{c^{k+1}\sigma_0} * I - g_{c^k\sigma_0} * I)(\mathbf{p}) \end{aligned} \quad (3.22)$$

où c est un facteur d'échelle. Pour e échelles par octave, il faut prendre $c = 2^{1/e}$ pour couvrir une octave. Une pyramide multi-résolution est habituellement construite pour une implémentation plus efficace. En effet, lorsqu'on double σ , la précision d'échantillonnage reste la même lorsque la résolution de l'image est diminuée de moitié. Cette pyramide est construite selon les étapes suivantes :

- La première image de la première octave est construite en doublant la résolution de l'image initiale par interpolation linéaire (pour contrecarrer les effets du lissage sur les détails de l'image). On suppose que la valeur de σ associée à cette image vaut $\sigma = 1$. Un lissage est alors effectué pour obtenir la valeur de σ_0 souhaitée.
- Les différentes images des différentes échelles d'une octave sont ensuite obtenues en lissant de nouveau l'image de l'échelle précédente. Pour obtenir la valeur de σ_{k+1} souhaitée, ce lissage doit être effectué avec un écart-type de $\sqrt{\sigma_{k+1}^2 - \sigma_k^2}$.
- Ensuite, l'image de l'octave courante calculée avec l'écart-type σ_k égal au double de celui de la première image de l'octave (l'antépénultième image lissée de l'octave courante), est sous-échantillonnée en prenant un pixel sur deux pour construire le prochain niveau de la pyramide (correspondant à l'octave suivante).
- Le processus est répété pour le nombre d'octaves souhaité, voir figure 3.5.

Les extrema dans l'espace échelle constituent les candidats qui sont « relocalisés » au sous-pixel. Pour cela, la fonction de réponse est interpolée localement par un développement limité de Taylor à l'ordre 2. Soit SI' la fonction qui donne la valeur de réponse pour une position et une échelle donnée :

$$\begin{aligned} \text{SI}' : \mathbb{R}^3 &\rightarrow \mathbb{R} \\ \begin{bmatrix} i \\ j \\ \sigma \end{bmatrix} &\mapsto (\text{SI}(\mathbf{p}))_{\sigma_k} \end{aligned} \quad (3.23)$$

où σ_k est l'indice du vecteur des réponses associé à l'échelle σ . Pour simplifier les notation, on pose $f = \text{SI}'([i \ j \ \sigma]^\top)$. En utilisant la formule de Taylor, on a :

$$\begin{aligned} \text{SI}' \left(\begin{bmatrix} i + \Delta_i \\ j + \Delta_j \\ \sigma + \Delta_\sigma \end{bmatrix} \right) &\approx f + f_i \Delta_i + f_j \Delta_j + f_\sigma \Delta_\sigma \\ &+ \frac{1}{2} \left(f_{ii} \Delta_i^2 + 2f_{ij} \Delta_i \Delta_j + 2f_{i\sigma} \Delta_i \Delta_\sigma \right. \\ &\left. + f_{jj} \Delta_j^2 + 2f_{j\sigma} \Delta_j \Delta_\sigma + f_{\sigma\sigma} \Delta_\sigma^2 \right) \end{aligned} \quad (3.24)$$

avec

$$\begin{aligned} f_i &= \frac{\partial f}{\partial i} & f_j &= \frac{\partial f}{\partial j} & f_\sigma &= \frac{\partial f}{\partial \sigma} \\ f_{ii} &= \frac{\partial^2 f}{\partial i^2} & f_{ij} &= \frac{\partial^2 f}{\partial i \partial j} & f_{i\sigma} &= \frac{\partial^2 f}{\partial i \partial \sigma} \\ f_{jj} &= \frac{\partial^2 f}{\partial j^2} & f_{j\sigma} &= \frac{\partial^2 f}{\partial j \partial \sigma} & f_{\sigma\sigma} &= \frac{\partial^2 f}{\partial \sigma \partial \sigma}. \end{aligned} \quad (3.25)$$

D'où, en dérivant selon Δ_i , Δ_j et Δ_σ :

$$\begin{cases} \frac{\partial}{\partial \Delta_i} \text{SI}' \left([i + \Delta_i \ j + \Delta_j \ \sigma + \Delta_\sigma]^\top \right) \approx f_i + f_{ii} \Delta_i + f_{ij} \Delta_j + f_{i\sigma} \Delta_\sigma \\ \frac{\partial}{\partial \Delta_j} \text{SI}' \left([i + \Delta_i \ j + \Delta_j \ \sigma + \Delta_\sigma]^\top \right) \approx f_j + f_{ij} \Delta_i + f_{jj} \Delta_j + f_{j\sigma} \Delta_\sigma \\ \frac{\partial}{\partial \Delta_\sigma} \text{SI}' \left([i + \Delta_i \ j + \Delta_j \ \sigma + \Delta_\sigma]^\top \right) \approx f_\sigma + f_{i\sigma} \Delta_i + f_{j\sigma} \Delta_j + f_{\sigma\sigma} \Delta_\sigma. \end{cases} \quad (3.26)$$

Il s'agit de chercher le vecteur de déplacement $[\Delta_i \ \Delta_j \ \Delta_\sigma]^\top$ qui donne le maximum de la fonction SI' , c'est-à-dire le vecteur qui annule ses dérivées partielles, d'où l'équation :

$$\begin{bmatrix} f_{ii} & f_{ij} & f_{i\sigma} \\ f_{ij} & f_{jj} & f_{j\sigma} \\ f_{i\sigma} & f_{j\sigma} & f_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \Delta_i \\ \Delta_j \\ \Delta_\sigma \end{bmatrix} = - \begin{bmatrix} f_i \\ f_j \\ f_\sigma \end{bmatrix} \quad (3.27)$$

et le vecteur d'ajustement de la position :

$$\begin{bmatrix} \Delta_i \\ \Delta_j \\ \Delta_\sigma \end{bmatrix} = - \begin{bmatrix} f_{ii} & f_{ij} & f_{i\sigma} \\ f_{ij} & f_{jj} & f_{j\sigma} \\ f_{i\sigma} & f_{j\sigma} & f_{\sigma\sigma} \end{bmatrix}^{-1} \begin{bmatrix} f_i \\ f_j \\ f_\sigma \end{bmatrix} \quad (3.28)$$

où les dérivées sont calculées par différences finies des valeurs des pixels voisins, voir figure 3.6.

De plus, les candidats situés dans des zones à faible contraste sont éliminés. Notons $f_{\max} = \text{SI}'([i + \Delta_i \ j + \Delta_j \ \sigma + \Delta_\sigma]^\top)$. En substituant (3.28) dans (3.24), les auteurs

de [Lowe 99] proposent la contrainte suivante :

$$|f_{\max}| = \left| f + \frac{1}{2} \begin{bmatrix} f_i & f_j & f_\sigma \end{bmatrix} \begin{bmatrix} \Delta_i \\ \Delta_j \\ \Delta_\sigma \end{bmatrix} \right| \geq t_c \quad (3.29)$$

où t_c est un seuil de contraste.

Enfin, les candidats situés sur des contours sont éliminés. La courbure principale dans la direction normale à un contour est forte et celle dans la direction du contour est faible. Les courbures principales étant proportionnelles aux valeurs propres de la matrice hessienne, seuls les points où le rapport entre la plus grande et la plus petite valeur propre est inférieur à un seuil t_r sont conservés. Ce rapport pouvant être exprimé en fonction de la trace et du déterminant de la matrice hessienne, la contrainte est donnée par :

$$\frac{\text{Tr}(\mathbf{H}_{\mathbf{p},\sigma})}{\text{Det}(\mathbf{H}_{\mathbf{p},\sigma})} = \frac{(f_{ii} + f_{jj})^2}{f_{ii}f_{jj} - f_{ij}^2} < \frac{(t_r + 1)^2}{t_r} \quad (3.30)$$

où $\mathbf{H}_{\mathbf{p},\sigma}$ désigne ici la matrice hessienne de SI' au point \mathbf{p} et à l'échelle σ . Les dérivées sont calculées par différences finies et t_r est un seuil.

- **Harris-Laplace** – (HL) [Mikolajczyk 04] – Les auteurs ont montré dans [Mikolajczyk 01] qu'avec le laplacien, si un point d'une image d'échelle σ_{i_1} présente un maxima à l'échelle σ_1 et si son correspondant dans une image d'échelle σ_{i_2} donne également un maxima à l'échelle σ_2 , alors le rapport $\sigma_{i_1}/\sigma_{i_2} = \sigma_1/\sigma_2$ est conservé plus souvent qu'avec la réponse de Harris. Ainsi, les auteurs proposent d'adapter Harris avec une recherche des extrema dans l'espace-échelle du laplacien. Deux étapes sont alors réalisées :

- ▶ *Calcul de la réponse de Harris à plusieurs échelles* – Les échelles sont définies par $\sigma_{I_k} = c^k \sigma_0$, où σ_D est fixé à $\sigma_D = \lambda \sigma_I$ avec $\lambda \in \mathbb{R}$ et où σ_D est l'échelle de différenciation et σ_I est l'échelle d'intégration, voir § 3.1.1. Les candidats sont obtenus par sélection des extrema locaux de la réponse.
- ▶ *Sélection des points d'intérêt dans l'espace échelle du laplacien* – Pour cette étape, les auteurs proposent un algorithme itératif appliqué à chaque candidat. Soient $[i_0 \ j_0 \ \sigma_0]^\top$ les coordonnées et l'échelle initiale du candidat considéré :
 - Recherche d'un extrema pour le candidat dans l'espace échelle du laplacien pour les échelles comprises dans l'intervalle $[c^{-1}\sigma_{I_k}, c\sigma_{I_k}]$ et où l'intervalle inter-échelle est précisé par $c_2 < c$. Si un extrema n'est pas trouvé alors ce candidat est rejeté. On note σ_{ex} l'échelle à laquelle cet extrema a été trouvé.
 - Détection de la position $[i_{k+1} \ j_{k+1}]^\top$ du maxima de la réponse de Harris le plus proche de $[i_k \ j_k]^\top$ à l'échelle σ_{ex} .
 - Reitérer si $\sigma_k \neq \sigma_{k+1}$ ou $[i_k \ j_k]^\top \neq [i_{k+1} \ j_{k+1}]^\top$.

Une version simplifiée et moins coûteuse en calculs de cette sélection a également été proposée. Parmi les points d'intérêt donnés par le détecteur de Harris initial à plusieurs

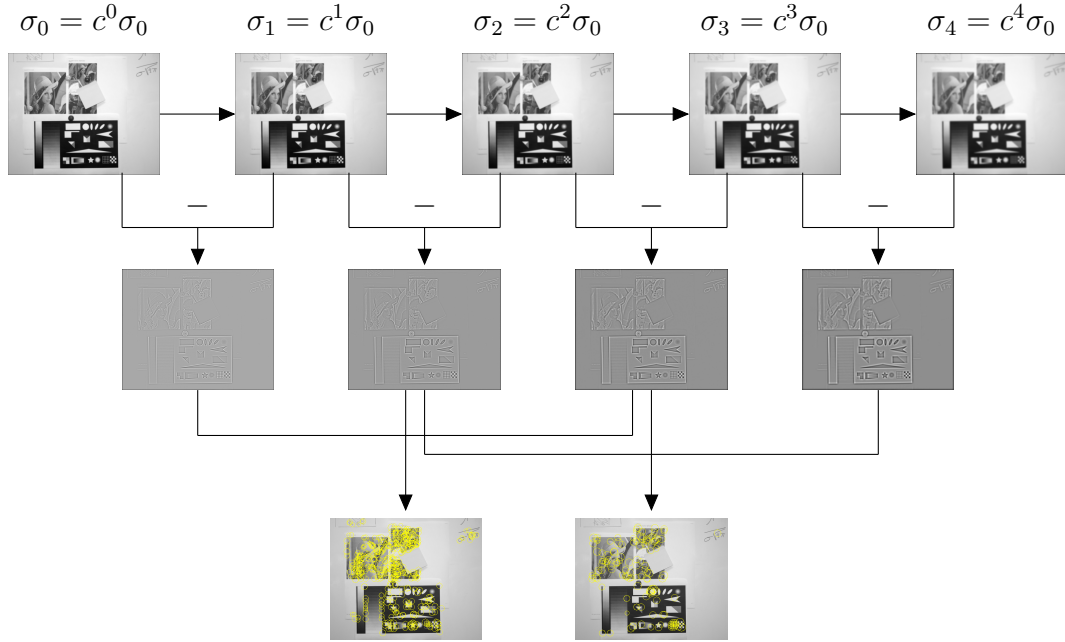


FIGURE 3.5: *SIFT – Différences de gaussiennes et sélection des extrema.* – Exemple de construction d'une octave avec 2 échelles, d'où $c = 2^{(1/2)} = \sqrt{2}$. La première image de la première octave est obtenue en effectuant un lissage et en doublant la résolution de l'image initiale. La première image des autres octaves sont obtenues en prenant un pixel sur deux de l'antépénultième image de l'octave précédente (car son échelle vaut $c^e \sigma_0 = 2^{1/e} \sigma_0 = 2\sigma_0$). À chaque octave, les images suivantes sont obtenues par lissage de l'image à l'échelle précédente. Enfin, les points d'intérêt candidats sont obtenus par sélection des extrema de ces différences dans l'espace échelle (pour chaque échelle, un candidat doit être un extrema selon les dimensions spatiales et par rapport à l'échelle inférieure et à l'échelle supérieure).

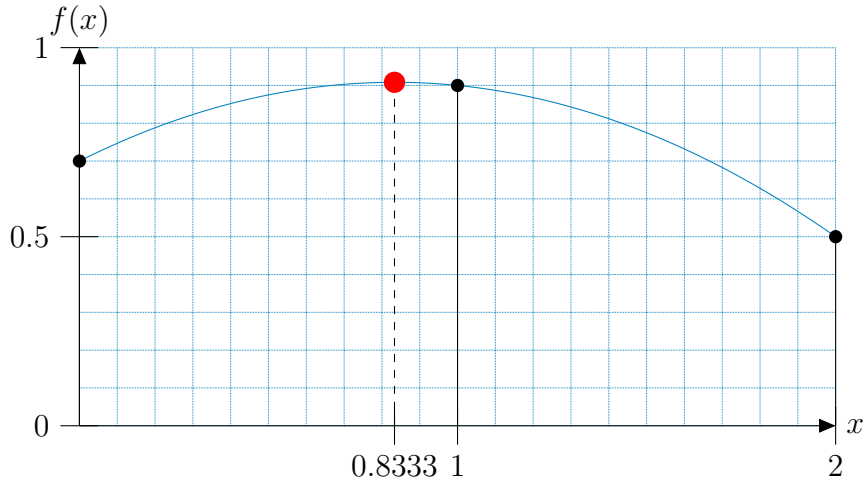


FIGURE 3.6: *Relocalisation.* – Sur cet exemple, la réponse $f(x)$ est donnée en 1D pour trois pixels $x = \{0, 1, 2\}$. La position relocalisée du candidat $x = 1$ est marquée d'un point rouge.

échelles, on ne conserve que ceux pour lesquels il existe un extrema dans l'espace échelle du laplacien et dont la réponse est supérieure à seuil donné.

REMARQUE : les points issus de ce détecteur sont utilisés notamment par le détecteur de régions Harris-Affine qui consiste à extraire des paramètres supplémentaires pour caractériser les régions au voisinage des points d'intérêt.

- **Hessian-Laplace** – (HEL) [Mikolajczyk 04]. Le principe est analogue à celui décrit pour Harris-Laplace, mais le détecteur de Harris est remplacé par le détecteur de Beaudet (déterminant de la matrice hessienne).
- **SURF** – (SF) [Bay 06] – La réponse est basée sur une approximation du déterminant de la matrice hessienne à plusieurs échelles à l'aide d'une implémentation astucieuse. La réponse est théoriquement donnée par :

$$\text{SF}(\mathbf{p}) = \begin{bmatrix} \vdots \\ (g_{\sigma_k} * I_{ii})(\mathbf{p})(g_{\sigma_k} * I_{jj})(\mathbf{p}) - (w(g_{\sigma_k} * I_{ij})(\mathbf{p}))^2 \\ \vdots \end{bmatrix} \quad (3.31)$$

où w est une constante de pondération fixée à $w = 0.9$.

Les auteurs ont remarqué qu'en pratique, le calcul des dérivées avec un masque de convolution constitue déjà une approximation. Toutefois, ils remarquent également que cette perte de précision n'empêche pas les détecteurs basés sur le laplacien de donner de bons résultats en terme de répétabilité et proposent d'utiliser directement des approximations moins coûteuses en calculs. Pour cela, ils proposent d'utiliser des masques

de dérivation au second ordre lissés par une gaussienne et tronqués de telle sorte qu'on puisse réaliser les calculs à l'aide d'images intégrales.

Image intégrale

L'image intégrale est définie par :

$$I_{\Sigma}(\mathbf{p}) = \sum_{k=0}^{k < i} \sum_{l=0}^{l < j} I\left(\begin{bmatrix} k \\ l \end{bmatrix}\right). \quad (3.32)$$

Elle permet de calculer facilement la somme de toutes les valeurs de I situées dans un rectangle délimité par les points de coordonnées \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3 et \mathbf{p}_4 par trois opérations arithmétiques (voir figure 3.7) :

$$\sum_{i=a}^b \sum_{j=c}^d I\left(\begin{bmatrix} i \\ j \end{bmatrix}\right) = I_{\Sigma}\left(\begin{bmatrix} b \\ d \end{bmatrix}\right) - I_{\Sigma}\left(\begin{bmatrix} a \\ d \end{bmatrix}\right) - I_{\Sigma}\left(\begin{bmatrix} b \\ c \end{bmatrix}\right) + I_{\Sigma}\left(\begin{bmatrix} a \\ c \end{bmatrix}\right). \quad (3.33)$$

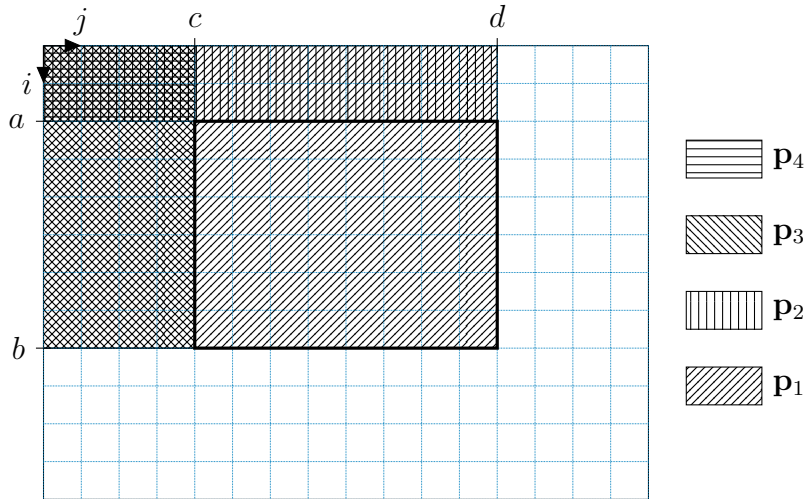


FIGURE 3.7: Image intégrale. – Calculer $I_{\Sigma}(\mathbf{p}_1) - I_{\Sigma}(\mathbf{p}_2) - I_{\Sigma}(\mathbf{p}_3) + I_{\Sigma}(\mathbf{p}_4)$, revient à calculer la somme des valeurs des pixels situés dans le rectangle marqué en gras sur l'image initiale.

Box filtering

Les masques utilisés pour calculer les dérivées sont construits de telle sorte qu'il ne faille calculer que des combinaisons de sommes de valeurs de pixels situés dans les différents rectangles (*box*). Ainsi, l'image intégrale peut être utilisée pour réduire le nombre d'opérations arithmétiques, voir figure 3.8. Les réponses aux différentes échelles sont obtenues en utilisant des masques de tailles différentes.

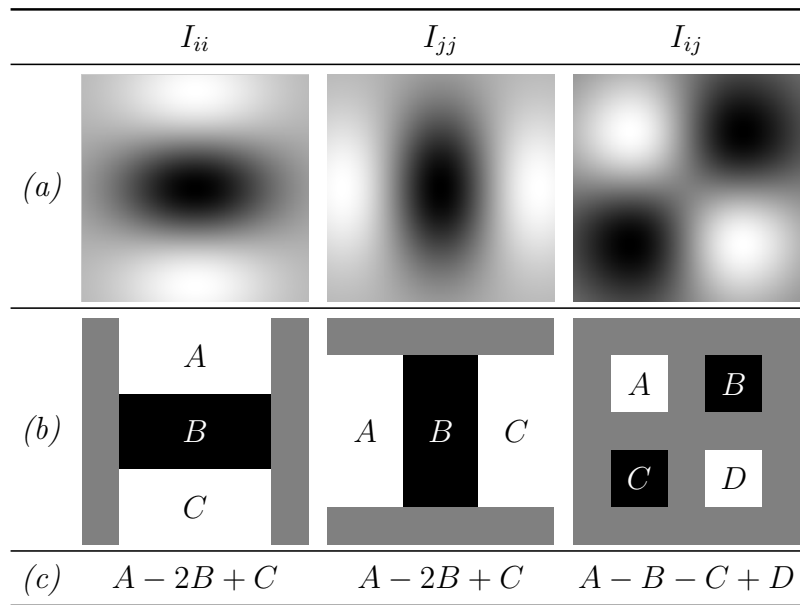


FIGURE 3.8: *Box filtering*. – La ligne (a) montre les images des masques de convolution pour calculer $g_\sigma * I_{ii}$, $g_\sigma * I_{jj}$ et $g_\sigma * I_{ij}$ pour un pixel \mathbf{p} . Plus le niveau de gris est élevé, plus la valeur du masque est élevée (ces valeurs pouvant être négatives, du gris représente 0). Les valeurs de ces masques sont approchées et tronquées pour obtenir les masques de la ligne (b). Ainsi, le calcul de la convolution peut être remplacé par l'opération indiquée en (c), où, par exemple, A est la somme des valeurs des pixels de l'image inclus dans le rectangle A lorsque le masque est centré sur \mathbf{p} . Cette somme peut être calculée rapidement à l'aide de l'image intégrale.

3.1.2 Famille morphologique

Les détecteurs de la famille morphologique s'appuient sur l'étude de la forme des régions constituées des pixels du voisinage ayant une valeur proche de celle du pixel considéré. Les détecteurs présentés ici sont à échelle fixe.

■ **SUSAN** (*Smallest Univalve Segment Assimilating Nucleus*) – (SU) [Smith 97] – La réponse s'appuie sur la surface des pixels de l'USAN, c'est-à-dire des pixels de valeurs similaires de celle du pixel considéré (ici appelé *nucleus*) dans un voisinage circulaire, voir figure 3.9. Ainsi, il est possible de distinguer les cas suivants :

- ▶ si la surface de l'USAN est proche de la surface du voisinage alors le pixel est situé dans une zone homogène ;
- ▶ si la surface de l'USAN est proche de la moitié de la surface du voisinage alors le pixel est situé sur un contour ;
- ▶ si la surface de l'USAN est plus petite que la moitié de la surface du voisinage alors le pixel est un coin ; c'est un point d'intérêt candidat.

La surface de l'USAN centrée sur le pixel de coordonnées \mathbf{p} est donnée par la fonction suivante :

$$s : \mathbb{N}^2 \rightarrow \mathbb{R}$$

$$\mathbf{p} \mapsto \sum_{k,l} \exp \left(\frac{I(\mathbf{p}) - I \left(\begin{bmatrix} i+k & j+l \end{bmatrix}^\top \right)}{t} \right)^6 \quad (3.34)$$

avec $(k, l) \in \mathbb{N}^2$ tel que $(k-i)^2 + (l-j)^2 < r^2$, et où r est le rayon du voisinage considéré et t est le paramètre de similarité. La valeur constante de la surface maximale est donnée par s_{\max} , le nombre de pixels du voisinage. La réponse pour un pixel est donnée par :

$$\text{SU}(\mathbf{p}) = \begin{cases} g - s(\mathbf{p}) & \text{si } s(\mathbf{p}) < g \\ 0 & \text{sinon} \end{cases} \quad (3.35)$$

où g est un seuil géométrique. En effet, le détecteur SUSAN peut également être utilisé pour détecter des contours. Pour détecter des points d'intérêt, on prend généralement $g \leq \frac{s_{\max}}{2}$.

Enfin, certains candidats sont rejetés à l'aide de deux contraintes :

- ▶ *Contrainte de centroïde* – Le centroïde de l'USAN doit être éloigné d'une distance d (en pixels) du *nucleus*. Les coordonnées $[i_c \ j_c]^\top$ du centroïde de l'USAN sont données par :

$$i_c = \frac{\sum_{k,l} (i+k) s(\mathbf{p})}{\sum_{k,l} s(\mathbf{p})} \quad \text{et} \quad j_c = \frac{\sum_{k,l} (j+l) s(\mathbf{p})}{\sum_{k,l} s(\mathbf{p})}. \quad (3.36)$$

■ 3 Sélection des germes initiaux pour la mise en correspondance par propagation

Pour qu'un pixel puisse être un point d'intérêt, il doit vérifier :

$$\sqrt{(i_c - i)^2 + (j_c - j)^2} > d \quad (3.37)$$

où d est un seuil.

- *Contrainte de contiguïté* – Tous les pixels du segment de droite défini par le *nucleus* et le centroïde de l'USAN doivent appartenir à l'USAN. On doit donc vérifier que pour tout pixel de coordonnées $[m \ n]^\top$ appartenant au segment défini par le *nucleus* et le centroïde :

$$\exp \left(\frac{(I(\mathbf{p}) - I([m \ n]^\top))^6}{t} \right) \leq t_u \quad (3.38)$$

où $t_u < 1$ est un seuil de similarité qui détermine si le pixel considéré appartient à l'USAN.

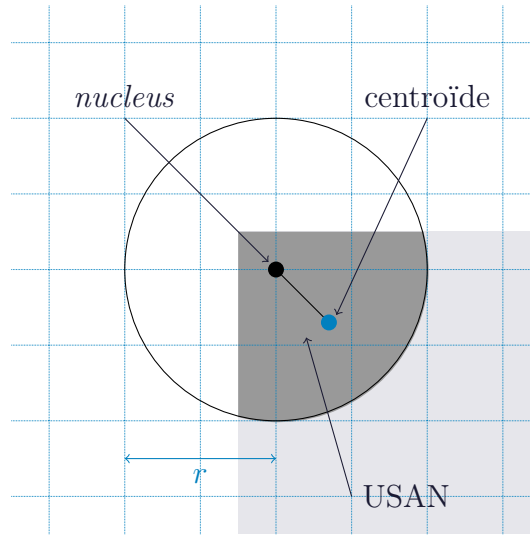


FIGURE 3.9: Détecteur SUSAN. – Les pixels dans les zones grises sont de même intensité que le *nucleus*. Les pixels dans le disque étant de même intensité que le *nucleus* constituent l'USAN (en gris foncé). Le centroïde de l'USAN est marqué par un point bleu. La contrainte de centroïde est vérifiée si la distance entre le *nucleus* et le centroïde est supérieure à un seuil. La contrainte de contiguïté est vérifiée si tous les pixels entre le *nucleus* et le centroïde appartiennent à l'USAN (ce qui est le cas sur cet exemple).

- **FAST** – (FA) [Rosten 06] – La réponse reflète la configuration des valeurs des pixels situés sur le périmètre d'un cercle centré sur le pixel considéré. Sur ce périmètre, constitué de n pixels, il doit y avoir $(3/4)n$ pixels contigus, soit plus clairs, soit plus sombres

que le pixel central. Un test rapide permet d'éliminer rapidement un candidat : il suffit de regarder les pixels situés au nord, à l'est, au sud, à l'ouest. Il doit y avoir au moins trois de ces quatre pixels soit plus clairs, soit plus sombres que le pixel considéré, voir figure 3.10. Une réponse est ensuite calculée pour les candidats sélectionnés afin d'effectuer la sélection des extrema locaux. La formule suivante est utilisée :

$$\text{FA}(\mathbf{p}) = \max \left(\sum_{\mathbf{q}_c \in \mathcal{P}_c} |I(\mathbf{q}_c) - I(\mathbf{p})| - t, \sum_{\mathbf{q}_s \in \mathcal{P}_s} |I(\mathbf{p}) - I(\mathbf{q}_s)| - t \right) \quad (3.39)$$

où t est un seuil de similarité et où \mathcal{P}_c est l'ensemble des coordonnées des pixels du périmètre du cercle plus clairs que le pixel considéré et \mathcal{P}_s désigne l'ensemble des coordonnées des pixels du périmètre plus sombres que le pixel considéré.

Les auteurs proposent de produire, à l'aide d'un processus d'apprentissage, un arbre de décision permettant de déterminer si un pixel est un point d'intérêt : les points d'intérêt sont d'abord détectés, en appliquant l'algorithme décrit précédemment, sur un ensemble d'images du domaine d'application. Cet arbre est ensuite traduit en programme informatique.

3.1.3 Famille entropique

Le calcul de la réponse est fondé sur l'entropie de la distribution des valeurs au voisinage des pixels. Le détecteur présenté ici est multi-échelle.

- **Kadir** – (KA) [Kadir 04] – La réponse repose sur l'entropie de la distribution des valeurs des pixels au voisinage du point considéré, c'est-à-dire sur le désordre, synonyme ici de « richesse » de l'information au voisinage du point considéré. Par exemple, si cette distribution est uniforme, cela signifie que le point étudié est situé dans une zone homogène et donc de peu d'intérêt. L'entropie est donnée par :

$$\begin{aligned} \mathcal{H} : \mathbb{N}^2 &\rightarrow \mathbb{R} \\ \mathbf{p} &\mapsto - \sum_{k,l} p(\mathbf{p}) \log(p(\mathbf{p})) \end{aligned} \quad (3.40)$$

avec $(k, l) \in \mathbb{N}^2$ et $(k - i)^2 + (l - j)^2 < r^2$ où r est le rayon du voisinage considéré et où la probabilité $p(\mathbf{p})$ est donnée par la distribution des valeurs des pixels du voisinage. En pratique, un histogramme des valeurs des pixels composé de w intervalles est utilisé pour estimer p . Une sélection des maxima locaux est ensuite appliquée sur la réponse suivante qui est calculée pour chaque candidat :

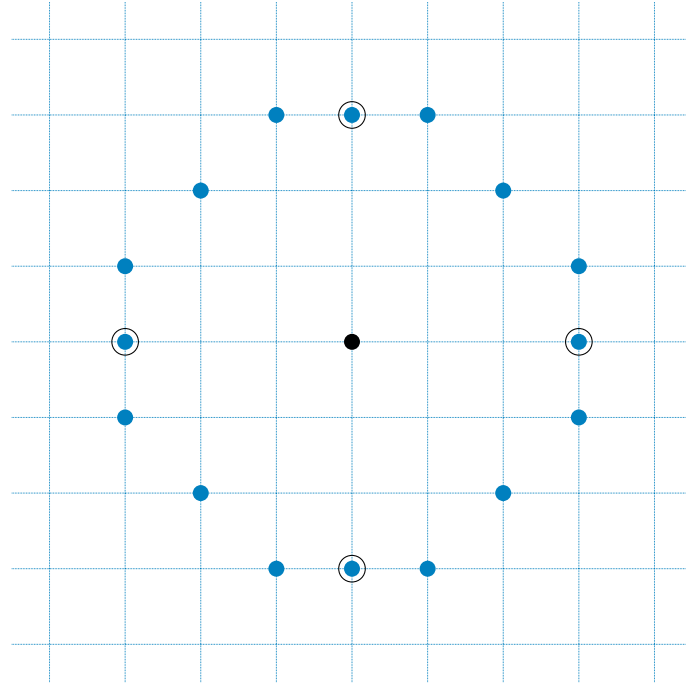


FIGURE 3.10: Détecteur FAST. – Ce détecteur étudie la configuration des valeurs des pixels sur le périmètre du cercle (discretisé par les points bleus) centré sur le pixel considéré (point noir). Les pixels situés au nord, à l'est, au sud et à l'ouest sont entourés en noir.

$$\text{KA}(\mathbf{p}) = \begin{bmatrix} \vdots \\ \mathcal{H}(\mathbf{p}) \frac{\sigma^2}{2\sigma-1} \sum_{k,l} \left| \frac{\partial p(\mathbf{p})}{\partial \sigma} \right| \\ \vdots \end{bmatrix}. \quad (3.41)$$

3.1.4 Prise en compte de la couleur

Les détecteurs présentés précédemment utilisent les niveaux de gris des pixels pour calculer leur réponse. Quelques travaux utilisent la couleur ou proposent des modèles de couleur pour remplacer les niveaux de gris. Dans [Gevers 97, Abdel-Hakim 06], les auteurs proposent deux modèles de couleur invariants à certaines transformations photométriques. Ces modèles sont utilisés dans le calcul de la réponse à la place des niveaux de gris. Leurs

expérimentations montrent que les points détectés sont plus stables en cas de changement de luminosité entre les deux images. Il apparaît également un gain en répétabilité d'environ 5% dans le meilleur des cas. Dans [Montesinos 98], le détecteur de Harris est adapté à la couleur en remplaçant, dans la matrice de Harris, les dérivées des niveaux de gris par une somme des dérivées de chaque composante du système RVB. Dans [Raoui 09], le détecteur de Harris est adapté à la couleur en remplaçant les dérivées des niveaux de gris par la somme des dérivées des composantes décorréliées du modèle couleur gaussien.

3.1.5 Synthèse des détecteurs

Les tables 3.1, 3.2 et 3.3 récapitulent la description des détecteurs de points d'intérêt, en particulier les étapes de post-traitement ainsi que les différents paramètres utilisés.

3.2 Appariement de points d'intérêt pour la sélection de germes

La mise en correspondance de points d'intérêt est un problème à part entière en analyse d'images car elle est utilisée dans différentes applications comme l'indexation d'images ou le suivi d'objets dans des séquences d'images. Nous pouvons distinguer deux catégories de méthodes :

- **Appariement de points d'intérêt par descripteurs robustes** – De nombreux descripteurs ont été proposés pour caractériser les points d'intérêt de manière invariante à différentes transformations géométriques (transformations affines) et photométriques (changements de luminosité). Une étude détaillée de ces descripteurs peut être trouvée dans [Mikolajczyk 05]. Ces transformations apparaissent dans des problèmes de reconnaissance d'objet, d'indexation ou de suivi où ces descripteurs sont parfaitement adaptés. Mais dans le cadre de la mise en correspondance de couples stéréoscopiques à « faible écartement », les déformations locales sont faibles. Ainsi, nous pensons que l'utilisation d'un descripteur robuste ne se justifie pas dans ce cas.
- **Appariement de points d'intérêt par corrélation** – Par conséquent, les pistes que nous avons explorées se concentrent sur une mise en correspondance de points d'intérêt par mesures de corrélation. Ces mesures sont identiques à celles présentées au § 2.2.3.1. Les problèmes les plus fréquemment rencontrés dans notre situation sont les occultations et les discontinuités de profondeur. Un détecteur de point d'intérêt, en particulier un détecteur de « coin », risque de détecter des points dans les zones proches des discontinuités de profondeur. Cela pose un problème pour les mesures de corrélation comme nous l'avons

■ 3 Sélection des germes initiaux pour la mise en correspondance par propagation

NOM	M-E	POST-TRAITEMENT	PARAMÈTRES
MO		■ Seuillage sur la réponse	■ n : échelle ■ t : seuil sur la réponse
HA		■ Seuillage sur la réponse	■ σ_D : échelle de différentiation ■ σ_I : échelle d'intégration ■ t : seuil sur la réponse
KR		■ Seuillage sur la réponse	■ n : échelle ■ t : seuil sur la réponse
BE		■ Seuillage sur la réponse	■ σ : échelle de différentiation ■ t : seuil sur la réponse
MIC		■ Seuillage sur la réponse	■ r : échelle telle que $ i + j = r$ dans le cas de l'interpolation carrée, $i^2 + j^2 = r$ dans le cas de l'interpolation circulaire ■ t : seuil sur la réponse
SI	✓	■ Seuillage sur la réponse ■ Relocalisation au sous-pixel ■ Élimination des candidats situés sur les contours	■ σ_0 : échelle initiale ■ o : nombre d'octaves ■ e : nombre d'échelles par octave ■ t_c : seuil sur la réponse ■ t_r : seuil sur les contours
HL	✓	■ Seuillage sur la réponse ■ Relocalisation au sous-pixel	■ σ_{I_0} : échelle d'intégration initiale ■ λ : facteur d'échelle tel que $\sigma_{D_k} = \lambda \sigma_{I_k}$ ■ e : nombre d'échelles souhaitées ■ c : intervalle entre deux échelles ■ t_h : seuil sur la réponse de Harris ■ c_2 : intervalle entre deux échelles pour la recherche des maxima (méthode non simplifiée)
HEL	✓	■ Seuillage sur la réponse ■ Relocalisation au sous-pixel	■ σ_0 : échelle initiale ■ e : nombre d'échelles souhaitées ■ c : intervalle entre deux échelles ■ t_h : seuil sur le déterminant de la matrice hessienne ■ c_2 : intervalle entre deux échelles pour la recherche des maxima (méthode non simplifiée)
SF	✓	■ Seuillage sur la réponse	■ n : échelle initiale ■ o : nombre d'octaves ■ e : nombre d'échelles par octave ■ t : seuil sur la réponse

TABLE 3.1: Détecteurs de la famille dérivée. – La colonne M-E indique si le détecteur est multi-échelle.

NOM	M-E	POST-TRAITEMENT	PARAMÈTRES
SU		<ul style="list-style-type: none"> ■ Seuillage de similarité ■ Seuillage géométrique ■ Contrainte du centroïde ■ Contrainte de contiguïté 	<ul style="list-style-type: none"> ■ r : échelle ■ t : seuil de similarité ■ g : seuil géométrique ■ d : seuil de distance pour la contrainte de centroïde ■ t_u : seuil de similarité pour la contrainte de contiguïté
FA		<ul style="list-style-type: none"> ■ Seuillage de similarité 	<ul style="list-style-type: none"> ■ r : échelle ■ t : seuil de similarité

TABLE 3.2: *Détecteurs de la famille morphologique.* – La colonne M-E indique si le détecteur est multi-échelle.

NOM	M-E	POST-TRAITEMENT	PARAMÈTRES
KA	✓	<ul style="list-style-type: none"> ■ Seuillage sur la réponse 	<ul style="list-style-type: none"> ■ n : échelle ■ t : seuil de similarité ■ w : nombre d'intervalles pour le calcul de l'histogramme

TABLE 3.3: *Détecteur de la famille entropique.* – La colonne M-E indique si le détecteur est multi-échelle.

expliqué au § 1.2. En revanche, la famille robuste constitue une bonne solution à ce problème.

La mise en correspondance des points d'intérêt diffère cependant d'une mise en correspondance dense au niveau de la liste des candidats à examiner. En effet, on suppose que le correspondant d'un point d'intérêt de l'image de référence est également un point d'intérêt de l'autre image (puisque les points d'intérêt issus d'un « bon » détecteur sont répétés). Les travaux présentés dans [Schmid 00] ont notamment permis de montrer que les points d'intérêt peuvent être sujets à un problème de localisation. Cela se traduit par le fait que l'hypothèse suivante n'est pas tout à fait exacte : le correspondant d'un point d'intérêt de l'image de gauche est également un point d'intérêt de l'image de droite, ou bien, l'un des voisins des points d'intérêt de l'image de droite. Nous pensons qu'une mise en correspondance des points d'intérêt de l'image de référence vers les points d'intérêt et leurs voisins de l'autre image est plus adaptée qu'une mise en correspondance de points d'intérêt vers points d'intérêt.

De plus, on attend des germes qu'ils soient fiables, c'est-à-dire que les appariements réalisés soient corrects, ce qu'on ne peut évidemment pas savoir *a priori*. Il ne paraît alors pas inutile

de vérifier des contraintes sur les appariements calculés. De nombreuses contraintes ont été proposées pour éliminer des appariements de points d'intérêt peu fiables, notamment à partir de la géométrie épipolaire, voir § 2.2.2.1. En outre, les contraintes présentées § 2.2.2.2 pour la mise en correspondance locale de pixels peuvent également s'appliquer. Une étude de l'utilisation de ces différentes contraintes pour la mise en correspondance de points d'intérêt peut être trouvée dans [Vincent 01, Zhou 02, Sur 10].

Malheureusement, ces contraintes, en plus de supprimer une partie des appariements erronés, peuvent également supprimer des appariements corrects. Par ailleurs, comme nous allons le voir dans le chapitre suivant, dans notre contexte de sélection de germes pour la mise en correspondance par propagation, il est important d'avoir un ensemble de germes bien répartis sur l'image et donc une cardinalité élevée de cet ensemble (or ces contraintes réduisent cette cardinalité). En outre, l'ajout de contraintes nécessite de déterminer les valeurs de nouveaux paramètres. La difficulté réside donc dans le compromis qui doit être fait et qui résulte du paramétrage.

3.3 Conclusion

Dans ce chapitre nous avons proposé un état de l'art des méthodes de sélection de germes pour la mise en correspondance par propagation. Ces germes, qui doivent être des appariements fiables, peuvent être obtenus par la mise en correspondance de points d'intérêt ou par sélection d'appariements fiables. Cette dernière approche nécessite de réaliser une mise en correspondance initiale dense dans laquelle les appariements qui vérifient un certain nombre de contraintes strictes sont conservés. La première approche ne nécessite que de mettre en correspondance des points d'intérêt. Il existe de nombreux détecteurs de points d'intérêt dont le comportement peut être réglé par plusieurs paramètres. De même, l'étape d'appariement de ces points d'intérêt comporte également différents paramètres. Une étude détaillée de la sélection des germes par appariements de points d'intérêt pour la mise en correspondance par propagation s'avère donc indispensable et fait l'objet du chapitre suivant.

4 Évaluation des méthodes de sélection de germes initiaux par appariement de points d'intérêt

DANS ce chapitre, nous évaluons les méthodes de sélection des germes initiaux par appariement de points d'intérêt. Ces germes servent de point de départ à la mise en correspondance par propagation. Les points d'intérêt sont obtenus en utilisant un détecteur parmi ceux décrits au chapitre précédent, voir § 3.1, page 76. Nous travaillons sur des couples d'images prises depuis des positions proches sur lesquels une rectification épipolaire a été appliquée, voir § 2.2.2.1, page 35. Lorsque les points d'intérêt sont situés près des zones de discontinuité de profondeur, il peut exister des différences sur une partie des voisinages des deux correspondants. Nous proposons de réaliser les appariements des points d'intérêt en utilisant des mesures de corrélation classiques et robustes aux niveaux des discontinuités de profondeur. Pour plus de lisibilité, nous appelons « couple détecteur-corrélation » un détecteur de points d'intérêt associé à une mesure de corrélation pour réaliser les appariements. Les déformations photométriques (changement de luminosité) et géométriques (rotation, changement d'échelle, changement perspective) entre les deux images des couples sont faibles. C'est la raison pour laquelle nous pensons que l'utilisation de descripteurs robustes à ces transformations pour l'appariement des points d'intérêt n'est pas particulièrement étudiée.

Nous commençons par définir les critères d'évaluation des germes dans le cadre de la mise en correspondance par propagation. Nous établissons ensuite un classement des meilleurs couples détecteur-corrélation satisfaisant au mieux ces critères. Enfin, nous proposons des

pistes pour améliorer les résultats obtenus tout en évaluant le gain ainsi réalisé.

Nos attentes sur les germes sont les suivantes :

- **Fiabilité** – Nous souhaitons que les germes initiaux soient corrects, c'est-à-dire que les appariements réalisés soient justes, afin d'éviter de propager trop d'erreurs dès le début du processus. Dans la pratique, il n'est pas possible de savoir *a priori* si les appariements calculés sont corrects. Mais on peut se poser par exemple la question suivante : sachant qu'un détecteur peut donner des points d'intérêt situés sur des coins, c'est-à-dire dans des zones proches des discontinuités de profondeur, une mesure de corrélation robuste permet-elle d'obtenir de meilleurs appariements qu'une mesure classique ? Par conséquent, l'évaluation proposée dans ce chapitre a pour objectif de comparer différents couples détecteur-corrélation pour déterminer quelles associations permettent d'obtenir, le plus souvent, les meilleurs résultats. La réponse n'est pas si évidente car de nombreux facteurs interviennent :
 - ▶ *La répétition des points d'intérêt détectés* – S'ils ne sont pas du tout répétés, il n'y aura alors aucune chance de trouver des appariements corrects et, surtout, on risque de trouver des appariements erronés. Toutefois, s'il existe un nombre élevé de points d'intérêt répétés à côté de points d'intérêt non répétés, on peut penser que la méthode d'appariement doit être capable d'écarter ces derniers.
 - ▶ *Le type de la mesure de corrélation utilisée* – Une mesure robuste est-elle plus performante qu'une mesure classique, notamment pour appairer les points d'intérêt situés au niveau des coins qui peuvent être dans les zones proches des discontinuités de profondeur ? Dans [Chambon 11], il a notamment été montré que les mesures robustes donnent généralement de meilleurs résultats dans les zones proches des discontinuités de profondeur mais que cela n'est pas forcément vrai en dehors de ces zones. Par conséquent, on voit bien qu'il existe un lien entre les points d'intérêt détectés et la mesure de corrélation utilisée pour les appairer qui peut influencer le résultat.
 - ▶ *Les contraintes sur les appariements* – Comme expliqué au chapitre précédent, des contraintes sur les appariements sont utilisées pour éliminer les moins probables. Elles rajoutent des paramètres et peuvent éliminer de mauvais appariements, mais également en supprimer des corrects. Par exemple, un appariement obtenu dans une zone proche d'une discontinuité de profondeur risque d'être éliminé par la contrainte de seuil s'il est obtenu avec une mesure de corrélation classique (car il peut avoir un score faible). En revanche, ce ne sera pas forcément le cas avec une mesure robuste, voir § 2.2.3.1, page 43.
- **Répartition** – Nous pensons qu'il est important que les germes soient « judicieusement » placés sur les images, c'est-à-dire de manière à ce qu'ils permettent, en se propageant, de trouver les appariements partout dans les images.

Les différentes étapes de sélection de germes par appariement de points d'intérêt font intervenir différents facteurs qui influencent le résultat et qui constituent les paramètres de la méthode :

- **Paramètres du détecteur de points d'intérêt** – Il en existe toujours au moins deux, l'échelle et le seuil sur la réponse. Les différents paramètres sont répertoriés § 3.1.1, page 81.
- **Paramètres de l'appariement des points d'intérêt** – Il s'agit notamment de la taille de la fenêtre de corrélation et de la zone de recherche. En effet, on peut considérer que le correspondant d'un point d'intérêt de l'image de référence se situe parmi les points d'intérêt de l'autre image ou des voisins. Ces paramètres sur l'appariement incluent également les paramètres sur les contraintes des appariements. Nous pensons que la contrainte de seuil sur le score de corrélation est indispensable, car elle permet d'éliminer des correspondants dont les voisinages ne se ressemblent pas. De plus, son coût en temps de calcul est très faible car elle ne nécessite qu'un seul test. Nous considérons que les autres contraintes sont optionnelles (et rajoutent des paramètres et des calculs plus complexes).

Afin de déterminer le paramétrage le plus approprié, une approche exhaustive qui consiste à faire varier indépendamment chaque paramètre et à observer le résultat obtenu est complexe en raison du nombre important de paramètres à étudier (au moins 5 paramètres par couple détecteur-corrélation). Face à ce problème, nous proposons d'appliquer à un échantillon de couples d'images pour lesquels la vérité terrain est connue, une méthode d'optimisation afin de déterminer, en fonction des critères souhaités, le paramétrage optimal des différents couples détecteur-corrélation évalués. Cette étude expérimentale permet de tirer des conclusions permettant de faciliter le paramétrage de l'étape de sélection des germes lorsque la vérité terrain n'est pas connue. Il s'agit de donner des indications sur les choix à faire pour trouver le meilleur compromis selon les différents objectifs souhaités pour le calcul de germes initiaux adaptés à une mise en correspondance par propagation.

Enfin, à partir d'une solution « de référence » ayant satisfait le mieux les critères souhaités, nous avons envisagé deux pistes pour améliorer les résultats. L'idée est de fusionner plusieurs ensembles de germes satisfaisant certains critères :

- La première piste vise à améliorer la répartition des germes en réalisant la fusion des meilleurs ensembles qui maximisent la fiabilité et la répartition. Ainsi, si les deux ensembles de départ ont une forte et même proportion de germes corrects par rapport au nombre total de germes, le risque que l'union des deux dégrade ce rapport est faible. En outre, la répartition ne peut qu'être améliorée si les deux ensembles sont complémentaires.
- La seconde piste cherche à améliorer la fiabilité des germes obtenus en fusionnant des ensembles qui maximisent uniquement la fiabilité. Cela permet d'obtenir des germes plus fiables (que si on avait aussi imposé de maximiser la répartition), mais moins nombreux et donc vraisemblablement moins bien répartis. Toutefois, en fusionnant les meilleurs ensembles, il est possible de retrouver au final une répartition équivalente à la solution « de référence » mais, cette fois, avec des germes plus fiables.

4.1 Critères d'évaluation

La fonction de disparité théorique est notée $d_{\text{th}}^{g \rightarrow d}$. L'erreur d'appariement est alors quantifiée par la fonction :

$$Err^{g \rightarrow d} : [0; h[\times [0; w[\rightarrow \mathbb{R}$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \mapsto \left| d_{\text{th}}^{g \rightarrow d} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) - d^{g \rightarrow d} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \right|. \quad (4.1)$$

Les vérités terrain sont données avec une précision au sous-pixel alors que les disparités que nous calculons sont données en pixel. On considère donc qu'un appariement est correct lorsque :

$$Err^{g \rightarrow d} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \leq 0.5 \quad (4.2)$$

c'est-à-dire lorsque la distance entre le correspondant trouvé et le correspondant théorique est inférieure à 0.5 pixel. Cette précision est illustrée par la figure 4.1 où le point bleu représente le correspondant théorique du pixel de référence considéré.

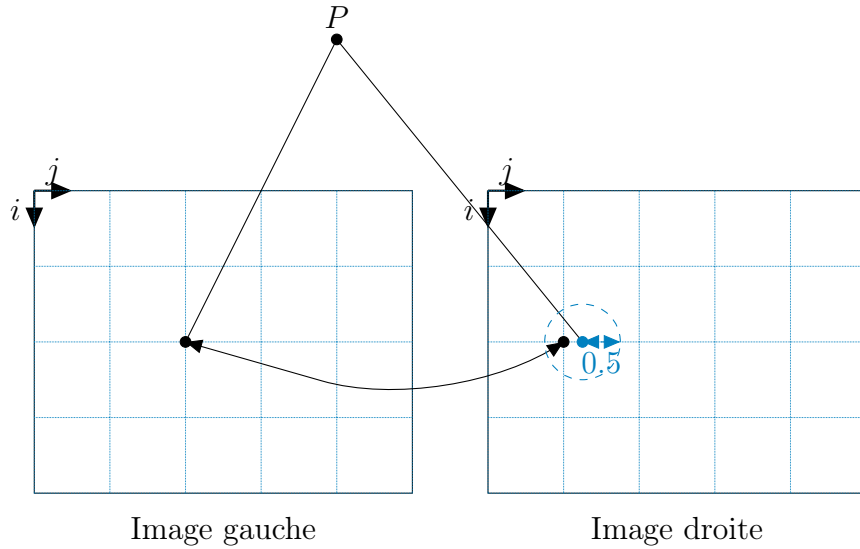


FIGURE 4.1: Précision de l'erreur d'appariement.

- **Taux de rappel** T_{ra} – Le taux de rappel évalue le rapport entre les germes corrects sur l'ensemble de germe calculés \mathcal{G} :

$$T_{ra}(\mathcal{G}) = \frac{\text{card} \left\{ \mathbf{g}_{i,j} \in \mathcal{G} \mid \text{Err} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \leq 0.5 \right\}}{\text{card } \mathcal{G}} \quad (4.3)$$

Ce critère est illustré par la figure 4.2. Sur cet exemple, les points d'intérêt sont marqués d'un point noir cerclé, les germes sont marqués d'une flèche représentant l'appariement entre deux correspondants. Nous avons ici, en violet (en haut), un appariement erroné, et en bleu (en bas), un appariement correct. Le taux de rappel est donc de 50%. Les points d'intérêt qui n'ont pu être mis en correspondance (car, par exemple, les appariements testés ne vérifient aucune des contraintes imposées) sont ignorés dans le calcul du taux de rappel.

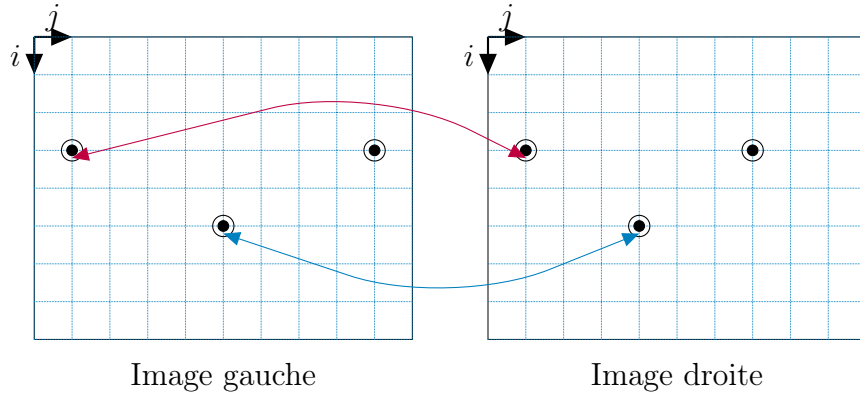


FIGURE 4.2: *Taux de rappel.*

- **Taux de répartition** $T_{ré}$ – Le taux de répartition permet d'évaluer la répartition des germes. Dans le cas de scènes qui présentent des discontinuités de profondeur, il faut qu'il existe au moins un germe par région de profondeurs homogènes pour avoir au moins une chance de réaliser la propagation dans ces régions. Pour évaluer la répartition, on propose de segmenter l'image en régions de couleurs homogènes (à l'aide de l'algorithme *mean-shift* présenté § 2.4, page 70) en faisant l'hypothèse que chaque région ne présente pas de discontinuité de profondeur et donc que cette segmentation est une sous-segmentation de la carte des profondeurs. La répartition est le rapport entre le nombre de régions qui contiennent au moins un germe correct et le nombre total de régions. Soit \mathcal{R} , l'ensemble des régions d'une image et \mathcal{I} , l'ensemble des pixels d'une

région, on a :

$$T_{ré}(\mathcal{G}, \mathcal{R}) = \frac{\text{card} \left\{ \mathcal{I} \in \mathcal{R} \mid \exists \mathbf{g}_{i,j} \in \mathcal{R}_{\mathcal{G}}^g \cap \mathcal{I} \text{ et } Err \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \leq 0.5 \right\}}{\text{card } \mathcal{R}} \quad (4.4)$$

où $\mathbf{g}_{i,j}$ est un pixel de référence d'un germe de l'ensemble \mathcal{G} . Une valeur faible du taux de répartition peut indiquer que peu de points d'intérêt ont été détectés dans une image présentant de nombreuses régions ou bien que la majorité des points d'intérêt détectés se situent dans une même région de l'image. Les régions sont obtenues en appliquant l'algorithme *mean-shift*, voir § 2.4, page 70. Nous avons utilisé le code fourni par <http://coewww.rutgers.edu/riul/research/code.html>. Nous avons fixé les paramètres de la segmentation en nous basant sur un critère visuel de telle sorte que nous obtenions de nombreuses régions (segmentation fine) afin que le taux de répartition calculé soit suffisamment représentatif de la répartition des germes sur la totalité de l'image. On souhaite, notamment, éviter qu'une segmentation moins fine ne donne un taux de répartition de 100% avec un nombre insuffisant de germes, ce qui serait inutile. Nous avons trouvé qu'avec les paramètres $h_r = h_s = 4$, nous obtenons des résultats satisfaisants sur les images que nous avons utilisées. Les différentes segmentations utilisées pour calculer le taux de répartition sont présentées dans la figure 4.3.

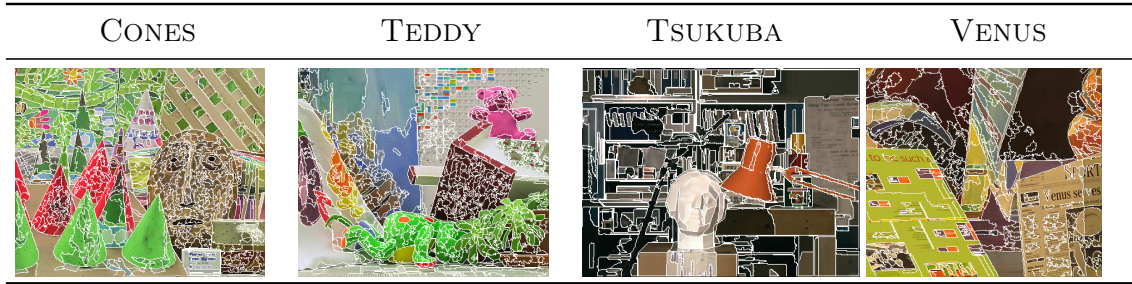


FIGURE 4.3: *Segmentations utilisées pour calculer le taux de répartition.*

Ce critère est illustré par la figure 4.4. Sur le premier exemple, trois régions contiennent chacune au moins un point d'intérêt (marqué d'un point noir entouré) : le taux de répartition est donc élevé. Sur le second exemple, tous les points d'intérêt sont dans la même région : le score de répartition est alors faible.

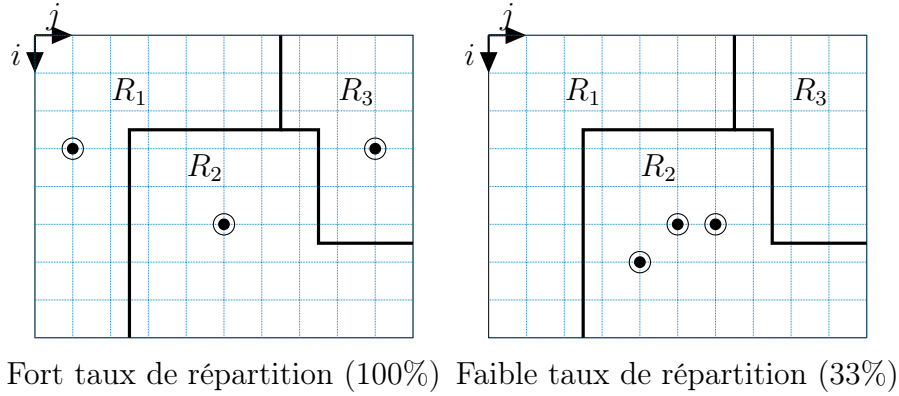


FIGURE 4.4: Taux de répartition.

4.2 Détermination expérimentale des paramètres de sélection des germes

Nous pensons que, dans une situation idéale, deux critères doivent être maximisés : le taux de rappel et le taux de répartition. En effet, il n'est pas utile d'avoir de nombreux germes corrects s'ils ne sont pas répartis de manière satisfaisante. Le risque est de ne pas pouvoir propager suffisamment les appariements sur toute l'image et donc d'avoir un résultat peu dense. Inversement, il ne sert à rien d'obtenir des germes bien répartis s'ils comportent trop d'erreurs. Le risque est de propager ces erreurs et d'obtenir un mauvais résultat. Par conséquent, nous proposons le compromis suivant :

$$C_A(\mathcal{G}, \mathcal{R}) = T_{ra}(\mathcal{G}) \times T_{ré}(\mathcal{G}, \mathcal{R}). \quad (4.5)$$

Par exemple, si nous trouvons 1000 germes corrects et 250 germes erronés, qui donnent un taux de rappel de $1000/(1000 + 250) = 80\%$, le taux de répartition permet de pondérer ce résultat. Si tous les germes sont situés dans une même région, ce qui n'est pas intéressant pour la propagation, alors ce taux sera faible, par exemple 10% , et donne un compromis de $80\% \times 10\% = 8\%$. En revanche, si les germes sont répartis sur les trois quarts des régions, le taux de répartition vaut 75% , et le compromis est égal à $80\% \times 75\% = 60\%$.

4.2.1 Protocole expérimental

Nous présentons dans cette section les images testées ainsi que l'ensemble des couples détecteur-corrélation. Nous rappelons les paramètres à déterminer et nous présentons également la technique d'optimisation employée.

4.2.1.1 Images testées

Les couples d'images que nous avons testés sont les quatre couples utilisés dans le protocole d'évaluation de Middlebury mis à disposition sur le site : vision.middlebury.edu/stereo/ et dont la description est détaillée dans [Scharstein 02, Scharstein 03, Scharstein 07, Hirschmüller 07]. Il s'agit de couples stéréoscopiques rectifiés à faible écartement qui présentent des scènes complexes avec différentes difficultés : zones homogènes, discontinuités de profondeur et motifs répétés, voir table 4.1. Les vérités terrain sont également mises à disposition sur le site. Pour l'image *Tsukuba*, la vérité terrain a été obtenue manuellement. Pour *Venus*, l'image est constituée de facettes planes et la vérité terrain est obtenue en estimant l'homographie associée à chaque plan. Pour les images *Cones* et *Teddy*, la vérité terrain a été obtenue à l'aide d'une technique de vision active par projection de lumière structurée. Une paire de caméras photographie la scène et un ou plusieurs projecteurs émettent une séquence de motifs lumineux qui permettent de calculer une signature unique pour chaque pixel, utilisée pour les mettre en correspondance de manière fiable. Le lecteur intéressé par les différentes techniques de lumière structurée pourra trouver plus d'informations dans [Fof 04].

REMARQUE : d'autres couples d'images sont également disponibles sur le site. Malheureusement, la détermination des paramètres par optimisation prend entre 1 à 2 semaines pour un seul couple. C'est la raison pour laquelle nous nous limitons à ces quatre couples qui servent de référence pour la comparaison des méthodes de mise en correspondance dense en stéréovision binoculaire.

4.2.1.2 Couples détecteur-corrélation testés

Les 70 couples détecteur-corrélation évalués sont obtenus en combinant 14 détecteurs avec 5 mesures de corrélation différentes. Les détecteurs utilisés ont été décrits § 3.1, page 76, et sont listés dans la table 4.2.

Les mesures de corrélation utilisées sont celles qui donnent les résultats les plus performants selon [Chambon 11]. Elles sont énumérées dans la table 4.3. Elles ont été décrites § 2.2.3.1, page 43 et pour l'évaluation, nous avons implémenté ces différents détecteurs en CUDA¹ afin d'utiliser les processeurs graphiques dans le but d'accélérer les calculs. Les temps de calcul sur une carte graphique NVIDIA FX4800 pour une détection et une mise en correspondance vont de quelques millisecondes à quelques secondes (≈ 20 s.) en fonction du détecteur, du nombre de points détectés et de la mesure de corrélation utilisée. Les détecteurs les plus lents sont les détecteurs multi-échelle. Les mesures de corrélation les plus lentes sont SMPD₂ et CENSUS.

1. www.nvidia.com/object/cuda_home_new.html

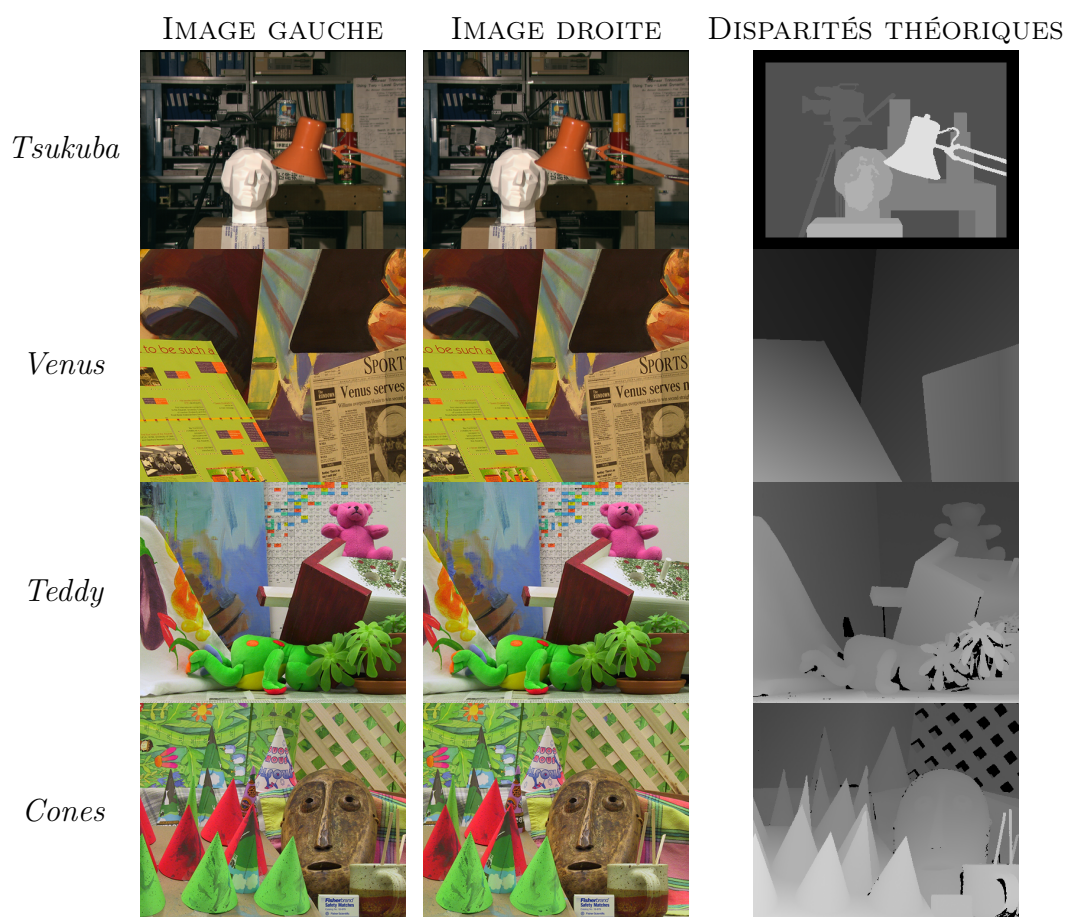


TABLE 4.1: Images testées. – Les couples d’images que nous avons testés sont ceux mis à disposition sur le site : vision.middlebury.edu/stereo/. La manière de calculer les disparité est donnée § 4.2.1.1.

ID.	DÉTECTEUR	ABR.
1	Moravec	MO
2	Harris, variante 1	HA.1
3	Harris, variante 2	HA.2
4	Harris, variante 3	HA.3
5	Kitchen et Rosenfeld	KR
6	Beaudet	BE
7	MIC	MIC
8	SUSAN	SU
9	FAST	FA
10	SIFT	SI
11	Harris-Laplace	HL
12	Hessian-Laplace	HEL
13	SURF	SF
14	Kadir	KA

TABLE 4.2: *Détecteurs de point d'intérêt comparés.* – La colonne ID. indique un numéro permettant d'identifier un détecteur, la colonne DÉTECTEUR donne le nom du détecteur et la colonne ABR. donne son abréviation.

ID.	CORRÉLATION
A	SAD
B	NCC
C	GC
D	SMPD ₂
E	CENSUS

TABLE 4.3: *Mesures de corrélation comparées.* – La colonne ID. indique une lettre permettant d'identifier une mesure, la colonne CORRÉLATION donne le sigle de la mesure.

4.2.1.3 Paramètres à déterminer

Paramètres du détecteur

Les paramètres des détecteurs évalués sont présentés § 3.1.1, page 81. Nous pouvons distinguer deux paramètres communs à tous les détecteurs :

- **Échelle** – Elle peut être exprimée sous trois formes différentes :

- ▶ $n \times n$ – Taille de la fenêtre prise en compte pour le calcul de la réponse.
- ▶ r – Rayon du disque pris en compte pour le calcul de la réponse.
- ▶ σ_D – Écart-type du filtre gaussien pour le calcul des dérivées.

Nous testons les échelles de $n \times n = 5 \times 5$ (ou $r = 2$, ou $\sigma = 1.25$) à $n \times n = 21 \times 21$ (ou $r = 10$, ou $\sigma = 10.5$). De même, pour les détecteurs multi-échelle nous limitons l'espace échelle à ces dimensions car nous pensons que pour la sélection des germes, il n'est pas nécessaire de chercher des points d'intérêt au delà. De plus, il n'existe pas de changement d'échelle entre l'image gauche et l'image droite des couples sur lesquels nous travaillons.

- **Seuil sur la réponse** – Il s'agit de la valeur de seuil sur la réponse du détecteur. Dans le but de faciliter l'interprétation des résultats, nous proposons une technique pour que ce seuil soit compris entre 0 et 1 pour tous les détecteurs.

Pseudo-normalisation du seuil du détecteur

Pour le seuil sur la réponse du détecteur, nous définissons de manière empirique les bornes de son intervalle de variation. La borne inférieure est telle que le seuillage n'élimine aucun point. La borne supérieure est telle que la quasi-totalité des points d'intérêt soit éliminée. Ces valeurs sont déterminées pour chaque échelle. Enfin, cet intervalle est ramené entre 0 et 1. Une valeur élevée (proche de 1) indique donc un seuillage très strict.

Paramètres de l'appariement

- **Taille de la fenêtre de corrélation** – Nous testons les dimensions $n \times n = 5 \times 5$ à $n \times n = 11 \times 11$. Une taille de fenêtre inférieure ou supérieure ne permet pas d'améliorer significativement les résultats selon nos expérimentations.
- **Taille de la zone de recherche des correspondants** – Nous testons les dimensions de $n \times n = 0 \times 0$ et $n \times n = 3 \times 3$. Autrement dit « sans » ou « avec » recherche des correspondants des points d'intérêt de l'image de référence au 8-voisinage des points d'intérêt de l'autre image.
- **Seuil sur le score de corrélation** – Nous utilisons une contrainte de seuil sur le score de corrélation pour éliminer les couples appariés dont les voisinages ne se ressemblent pas beaucoup. Comme pour le seuil sur la réponse du détecteur, nous proposons une technique pour que ce seuil soit compris entre 0 et 1 pour toutes les mesures de corrélation.

Pseudo-normalisation du paramètre de seuil de l'appariement

Nous définissons de manière empirique un intervalle de variation des scores de chaque mesure de corrélation. La borne inférieure est telle que le seuillage n'élimine aucun appa-

riement, la borne supérieure est telle que la quasi-totalité des appariements soit éliminée. Enfin, cet intervalle est ramené entre 0 et 1. Une valeur élevée (proche de 1) indique donc un seuillage très strict.

Autres contraintes sans paramètre sur l'appariement

- **Contrainte de symétrie** – Nous imposons la contrainte de symétrie. Les appariements doivent être identiques quelle que soit l'image de référence (gauche ou droite), voir § 2.2.2.2, page 37.
- **Contrainte d'ambiguïté-imprécision** – Nous proposons d'imposer également une contrainte d'ambiguïté-imprécision sans paramètre ajouté et dérivée de celles présentées § 2.2.2.2, page 37. Pour qu'un appariement soit accepté, il ne doit y avoir qu'un seul candidat tel que la contrainte de seuil soit vérifiée. En effet, les erreurs d'appariement sur les points d'intérêt (ou leur voisins) peuvent être dues au fait que, parmi l'ensemble des candidats, il en existe plusieurs dont les voisinages se ressemblent (en cas de motif répété par exemple). Ainsi, cette contrainte ne conserve que les appariements pour lesquels il n'existe qu'un seul candidat potentiel à la mise en correspondance.

4.2.1.4 Détermination des paramètres par optimisation

Compte tenu du nombre important de configurations possibles², une approche exhaustive consistant à tester toutes les solutions n'est pas apparue évidente à mettre en place même si faire varier tous les paramètres un par un permet d'observer leur influence sur le résultat. Nous proposons une approche permettant d'obtenir par optimisation les valeurs des différents paramètres qui satisfont nos critères, puis d'observer ces résultats pour en tirer des indications sur le paramétrage à effectuer. Pour cela, nous avons utilisé l'algorithme du recuit simulé que nous avons évoqué § 2.3.3, page 63. Il s'agit d'un algorithme itératif qui fait varier les différents paramètres et accepte, ou pas, selon une méthode décrite en détails ci-après, la nouvelle configuration obtenue. Cet algorithme converge vers une solution proche du maximum global du critère étudié [Kirkpatrick 83, Kirkpatrick 84]. Les grandes lignes de l'algorithme sont données pour le calcul des disparités dans le cadre d'une mise en correspondance globale § 2.3.3, page 66. Il peut également s'appliquer à notre problème en définissant :

- **Une description du système** – Il est donné par la liste des germes calculés avec le couple détecteur-corrélation et les paramètres donnés à une itération t .
- **Un déplacement (« move ») aléatoire** – Il faut définir comment passer d'une confi-

2. L'espace des solutions est au minimum de dimension 5 (échelle du détecteur, seuil sur la réponse du détecteur, dimensions de la fenêtre de corrélation, seuil sur la mesure de corrélation, taille de la zone de recherche des appariements) et, pour certaines dimensions, les valeurs possibles peuvent être réelles (en particulier pour les seuils), ce qui constitue un espace de recherche en théorie infini.

guration du système à une autre. Pour cela, un des paramètres du problème est tiré au sort (tirage selon une loi uniforme). Ensuite, la nouvelle valeur attribuée à ce paramètre est tirée au sort parmi l'ensemble des valeurs possibles pour ce paramètre, voir § 4.2.1.3. Pour chaque paramètre non discret, par exemple un seuil, nous avons au préalable défini de manière empirique un intervalle de valeurs possibles puis discrétisé de façon uniforme cet intervalle.

- **Critère à maximiser** – Elle est donnée par le compromis A décrit par l'équation (4.5) que l'on souhaite maximiser.
- **Une fonction de recuit** – Il s'agit de la fonction qui permet la décroissance de la température. Lorsqu'on tire au sort une nouvelle configuration, soit elle améliore le critère courant et elle est acceptée, soit elle ne l'améliore pas, et, dans ce cas, elle est acceptée avec une probabilité qui est d'autant plus grande que la température courante est élevée. Au fur et à mesure du recuit, la température doit baisser afin de converger vers le maximum global. En pratique, on arrête lorsque la température est trop basse, l'acceptation de nouvelles solutions étant alors moins probable, ou lorsqu'il n'y a plus de nouvelle solution acceptée.

La méthode d'optimisation par recuit simulé est décrite par l'algorithme 4.1.

4.2.2 Résultats

Nous commençons par présenter le classement des couples détecteur-corrélation selon le rang de compromis A, puis nous détaillons le paramétrage permettant d'obtenir ces résultats. Enfin, nous recalculons les germes à partir d'un paramétrage « optimal » commun à toutes ces images.

4.2.2.1 Classement des couples détecteur-corrélation selon le rang de compromis A

La figure 4.6, présente le classement des couples détecteur-corrélation triés par rang moyen selon le meilleur compromis A déterminé pour chaque couple d'images avec les paramètres obtenus par optimisation. Le rang moyen est calculé en faisant la moyenne des rangs, c'est-à-dire la moyenne des positions dans les classements par compromis A de chaque couple détecteur-corrélation, sur l'ensemble des couples d'images testés. Les 10 meilleurs couples détecteur-corrélation sont donnés dans la table 4.4, page 117.

On constate que l'écart-type sur le rang moyen varie entre 1.3 et 5.7 pour les 5 premiers couples détecteur-corrélation. Autrement dit, ces 5 couples se retrouvent tous dans les 10 meilleurs couples détecteur-corrélation pour chaque couple d'images testé. On peut donc dire que, globalement, ces couples donnent le plus souvent les meilleurs résultats sur le type d'images testé.

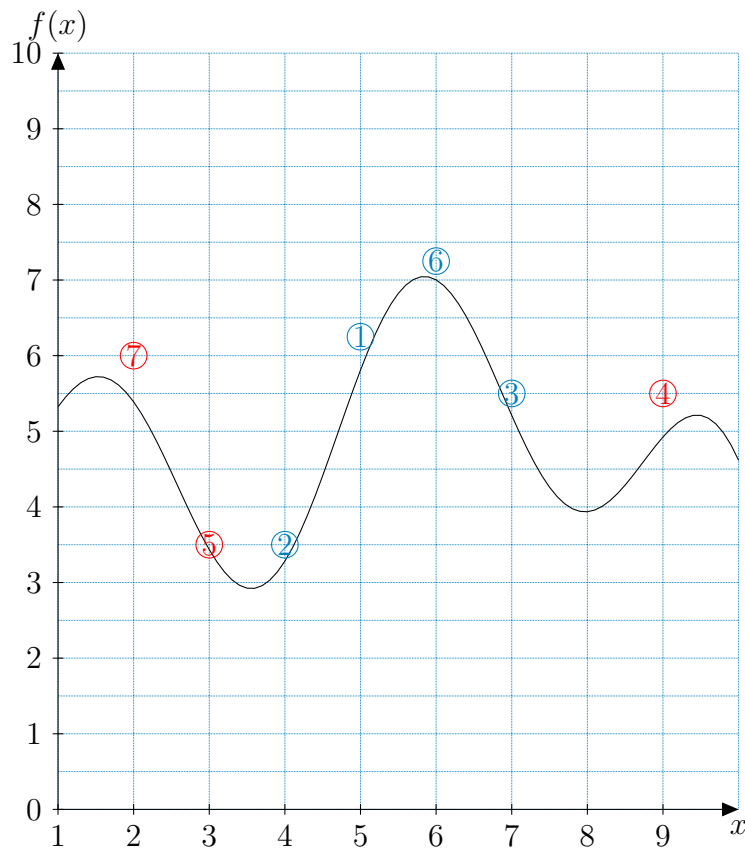


FIGURE 4.5: Recuit simulé. – Cette figure présente un exemple de recuit pour une fonction 1D f à maximiser. Chaque configuration x tirée au sort à une itération t est cerclée de bleu si elle est acceptée, de rouge si elle est refusée. Par exemple pour $t = 1$, on a $x = 5$ et, pour $t = 2$, $x = 4$. Au début du recuit, la température est encore haute et donc la solution marquée 2 est acceptée bien qu'elle donne une valeur de la fonction inférieure à celle de la configuration précédente. Au fur et à mesure des itérations, la température baisse et de moins en moins de configurations faisant baisser la valeur de la fonction sont acceptées. Ainsi, si on ne regarde que les solutions acceptées (en bleu), on remarque qu'elles convergent vers le maximum global (ici autour de $x = 6$).

```

1:  $\mathcal{P}$  : ensemble de paramètres à déterminer
2:  $\mathcal{P}_0$  : état initial de l'ensemble  $\mathcal{P}$ 
3:  $\mathcal{P}_t$  : état de l'ensemble  $\mathcal{P}$  à l'itération  $t$ 
4:  $\mathcal{P}_0 \leftarrow$  chaque paramètre  $p \in \mathcal{P}$  reçoit une valeur quelconque comprise dans un inter-
   valle discret
5:  $\mathcal{P} \leftarrow \mathcal{P}_0$ 
6:  $t \leftarrow 0$ 
7:  $T_0 \leftarrow$  température initiale
8: répéter
9:    $\mathcal{P}_t \leftarrow$  un paramètre  $p \in \mathcal{P}$  est tiré au sort et reçoit une valeur également tirée au
     sort comprise dans l'intervalle discret
10:   $T_t \leftarrow f_d(T_{t-1})$ 
11:  calculer la variation  $\Delta C = C(\mathcal{P}_t) - C(\mathcal{P}_{t-1})$ 
12:  si  $\Delta C > 0$  alors
13:    accepter le changement
14:  sinon
15:    accepter le changement avec une probabilité fonction de la température courante
16:  fin si
17:   $t \leftarrow t + 1$ 
18: jusqu'à convergence
19: retourne  $\mathcal{P}_t$ 

```

ALGORITHME 4.1: *Recuit simulé pour l'estimation des paramètres de la sélection des germes.*

Parmi les 10 meilleurs couples détecteur-corrélation, on retrouve les détecteurs :

- FAST (3 fois) ;
- KITCHEN (3 fois) ;
- SUSAN (2 fois) ;
- MIC (1 fois) ;
- HARRIS1 (1 fois).

On retrouve donc les deux détecteurs de la famille morphologique et d'autres détecteurs de la famille dérivée. Tous ces détecteurs sont à échelle fixe et détectent des points d'intérêt de type « coin ».

Parmi les dix meilleurs couples détecteur-corrélation, on retrouve les mesures de corrélation :

- CENSUS (5 fois) ;
- GC (3 fois) ;
- SAD (2 fois).

On remarque que sur les cinq premiers couples détecteur-corrélation, quatre utilisent la mesure de corrélation CENSUS. Il s'agit d'une mesure de la famille non-paramétrique qui s'avère robuste aux discontinuités de profondeur [Chambon 11]. Cela permet notamment de répondre à notre question concernant l'utilité d'une mesure de corrélation robuste pour l'appariement de points d'intérêt dont certains peuvent être situés sur les contours des objets et donc près des zones de discontinuité de profondeur. Il semble que CENSUS réponde à ces exigences. Les mesures GC et SAD donnent également de bons compromis. Elles présentent notamment l'avantage d'être beaucoup plus rapides à calculer que CENSUS.

4.2.2.2 Paramètres médians optimaux des couples détecteur-corrélation pour le compromis A

La table 4.5 donne les paramètres optimaux qui ont permis d'obtenir la valeur maximale de compromis A selon le recuit simulé pour les quatre couples d'images testés avec le couple détecteur-corrélation FAST-CENSUS. Bien entendu, en pratique, la vérité terrain n'étant pas connue, la méthode proposée pour déterminer les paramètres optimaux d'un couple détecteur-corrélation ne peut s'appliquer. L'objectif de l'étude proposée est de donner des indications sur les choix à faire concernant le paramétrage des couples détecteur-corrélation. On constate que les paramètres sont souvent les mêmes, ou qu'ils ont des valeurs très proches, au moins trois fois sur quatre, pour les différents couples d'images. Par conséquent, pour chaque couple détecteur-corrélation, on fixe dorénavant les paramètres à la valeur médiane donnée par les couples d'images testés. La table 4.6 indique la médiane des paramètres des 10 premiers couples détecteur-corrélation selon le compromis A obtenu pour tous les couples d'images. Nous avons également estimé l'écart-type de manière robuste [Huber 81]. Pour un ensemble de n valeurs x_i ($i \in [1; n]$), on prend $1.4826 \text{ med } |x_i - \text{med}(x_i)|$.

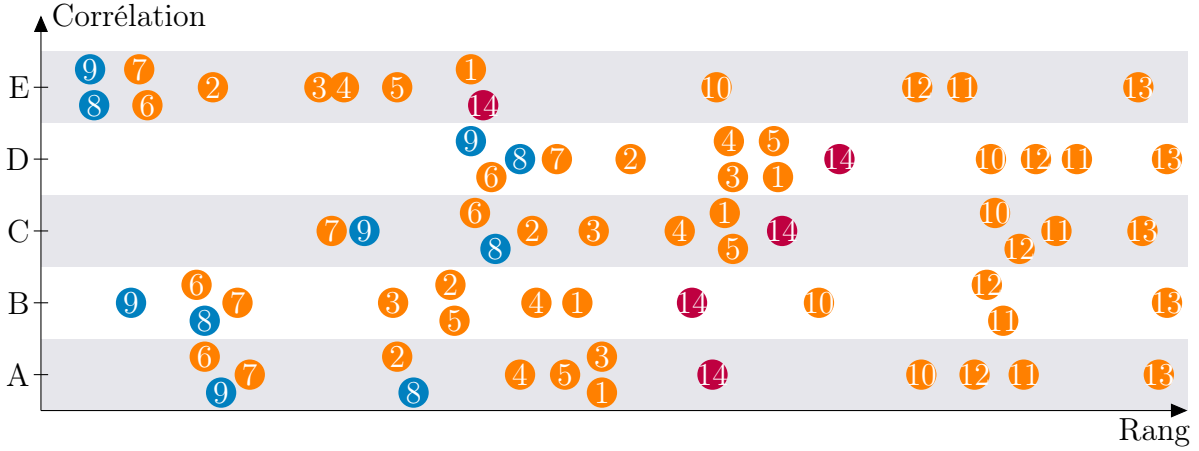


FIGURE 4.6: Classement des couples détecteur-corrélation par compromis A. – Cette figure présente le classement des rangs moyens des couples détecteur-corrélation par compromis A. Chaque pastille représente un détecteur indiqué par un numéro, voir § 4.2.1.2, page 108. Les pastilles orange correspondent aux détecteurs de la famille dérivée, les bleues à ceux de la famille morphologique et les violettes à celui de la famille entropique. La position en abscisse de chaque pastille indique son rang moyen selon le compromis A. La position en ordonnée indique la mesure de corrélation associée, voir § 4.2.1.2. On constate que le meilleur couple détecteur-corrélation selon le compromis A est 9E (pastille la plus à gauche) qui correspond au couple FAST-CENSUS.

COUPLE	RANG	C_A (%)	T_{ra} (%)	T_r (%)
FA-CENSUS	3.0 ± 2.3	44.8 ± 16.9	82.7 ± 11.9	53.3 ± 15.3
SU-CENSUS	3.3 ± 3.3	47.1 ± 16.8	80.2 ± 10.5	56.8 ± 14.4
FA-GC	5.5 ± 1.1	40.4 ± 14.2	77.2 ± 12.3	51.1 ± 11.5
MIC-CENSUS	6.0 ± 4.3	39.8 ± 12.3	80.5 ± 11.1	48.8 ± 10.9
KR-CENSUS	6.5 ± 5.0	39.4 ± 9.6	79.8 ± 10.8	48.7 ± 7.0
KR-GC	9.5 ± 3.2	36.9 ± 12.7	71.4 ± 11.3	50.2 ± 10.3
KR-SAD	10.0 ± 2.7	36.4 ± 13.5	69.7 ± 10.9	50.5 ± 11.5
SU-GC	10.0 ± 7.9	38.5 ± 13.1	75.4 ± 7.9	49.7 ± 13.3
HA.1-CENSUS	10.5 ± 3.9	35.9 ± 11.1	81.8 ± 9.8	43.0 ± 8.9
FA-SAD	11.0 ± 5.1	37.4 ± 16.3	76.4 ± 14.0	47.1 ± 13.5

TABLE 4.4: Classement des couples détecteur-corrélation par rang de compromis A. – Ce tableau présente les moyennes des rangs, suivies de leurs écarts-types, sur les quatre couples d'images testés. Nous présentons uniquement les 10 premiers couples détecteur-corrélation.

DÉTECTEUR			APPARIEMENT		
IMAGES	ÉCHELLE	SEUIL	TAILLE FENÊTRE	SEUIL	VOISINAGE
<i>Tsukuba</i>	$r = 2$	0.01	11×11	0.75	✓
<i>Venus</i>	$r = 2$	0.03	11×11	0.75	✓
<i>Teddy</i>	$r = 2$	0.01	11×11	0.717	✓
<i>Cones</i>	$r = 2$	0	11×11	0.667	✓

TABLE 4.5: *Paramètres optimaux pour le compromis A avec FAST-CENSUS.* – Cette table donne, pour le couple détecteur-corrélation FAST-CENSUS, les valeurs sur tous les couples d'images des paramètres suivants : l'échelle et le seuil du détecteur, la taille de la fenêtre ainsi que le seuil pour la mesure de corrélation. Une marque dans la dernière colonne indique si la recherche des correspondants d'un point d'intérêt de l'image de référence doit être effectuée au voisinage des points d'intérêt de l'autre image. Les paramètres donnés ici sont pseudo-normalisés selon la méthode décrite § 4.2.1.3.

DÉTECTEUR			APPARIEMENT		
MÉTHODE	ÉCHELLE	SEUIL	TAILLE FENÊTRE	SEUIL	VOISINAGE
FA-CENSUS	$r=2$	0.01	11×11	0.734	✓
SU-CENSUS	$r=2$	0	11×11	0.742	✓
FA-GC	$r=2$	0	11×11	0.49	✓
MIC-CENSUS	$n \times n=7 \times 7$	0.02	9×9	0.808	✓
KR-CENSUS	$n \times n=7 \times 7$	0	11×11	0.775	✓
KR-GC	$n \times n=5 \times 5$	0	9×9	0.5	✓
KR-SAD	$n \times n=5 \times 5$	0	11×11	0.88	✓
SU-GC	$r=2$	0	9×9	0.585	✓
HA.1-CENSUS	$\sigma_D=1.25$	0	9×9	0.775	✓
FA-SAD	$r=2$	0.02	9×9	0.89	✓

TABLE 4.6: *Paramètres médians optimaux pour le compromis A.* – Cette table donne les valeurs des paramètres suivants : l'échelle et le seuil du détecteur, la taille de la fenêtre ainsi que le seuil pour la mesure de corrélation. Une marque dans la dernière colonne indique si la recherche des correspondants d'un point d'intérêt de l'image de référence doit être effectuée au voisinage des points d'intérêt de l'autre image. Les paramètres donnés ici sont pseudo-normalisés selon la méthode décrite § 4.2.1.3.

Paramètres du détecteur

- **Échelle** – On constate que les échelles des détecteurs à échelle fixe sont petites. La valeur médiane de l'échelle de chaque détecteur à échelle fixe est donnée dans la table 4.7.

DÉTECTEUR	ÉCHELLE MÉDIANE
MORAVEC	$n \times n = 5 \times 5 \pm 0$
HARRIS1	$\sigma_D = 1.25 \pm 0$
HARRIS2	$\sigma_D = 1.25 \pm 0$
HARRIS3	$\sigma_D = 1.25 \pm 0$
KITCHEN	$n \times n = 5 \times 5 \pm 0$
BEAUDET	$n \times n = 7 \times 7 \pm 0$
MIC	$n \times n = 5 \times 5 \pm 0$
SUSAN	$r = 2 \pm 0$
FAST	$r = 2 \pm 0$

TABLE 4.7: *Échelles médianes des détecteurs à échelle fixe.*

Le fait de s'intéresser aux détails de l'image permet d'obtenir davantage de points d'intérêt que si on s'intéressait aux grandes structures et donc d'augmenter les chances d'obtenir une meilleure répartition des germes.

- **Seuil sur la réponse** – On remarque que le seuil sur la réponse du détecteur est très faible voire nul. Les seuils médians sont donnés pour chaque détecteur dans la table 4.8. Ces résultats sont donnés à titre indicatif car les intervalles calculés pour recadrer entre 0 et 1 les seuils des mesures non normalisées ont été définis de manière empirique, voir § 4.2.1.3. Cependant, ces résultats permettent de constater qu'il est préférable de ne pas contraindre le détecteur afin d'avoir le plus de points d'intérêt possible. C'est l'appariement qui a donc la « responsabilité » d'écarter les points d'intérêt qui ne sont pas répétés.

Paramètres de l'appariement

Nous donnons ici les résultats pour les paramètres communs à tous les couples détecteur-corrélation.

- **Taille de la fenêtre de corrélation** – Les dimensions médianes de la fenêtre de corrélation de chaque mesure utilisée sont données par la table 4.9. On note donc que les tailles de la fenêtre de corrélation les plus adaptées sont 9×9 et 11×11 . Le fait

DÉTECTEUR	SEUIL MÉDIAN
MORAVEC	0.01 ± 0
HARRIS1	0.0 ± 0.01
HARRIS2	0.0 ± 0
HARRIS3	0.0 ± 0
KITCHEN	0.0 ± 0
BEAUDET	0.0 ± 0
MIC	0.03 ± 0.01
SUSAN	0.0 ± 0
FAST	0.02 ± 0.01
SIFT	0.05 ± 0
HARRIS-LAPLACE	0.38 ± 0.16
HESSIAN-LAPLACE	0.01 ± 0
SURF	0.02 ± 0
KADIR	0.32 ± 0.02

TABLE 4.8: *Seuils médians des détecteurs.*

DÉTECTEUR	TAILLE MÉDIANE
CENSUS	$11 \times 11 \pm 0$
GC	$9 \times 9 \pm 2$
SAD	$9 \times 9 \pm 0$
SMPD ₂	$9 \times 9 \pm 2$
NCC	$10 \times 10 \pm 2$

TABLE 4.9: *Tailles médianes des fenêtres de corrélation.*

de prendre en compte davantage de voisins dans cette fenêtre augmente la fiabilité de mise en correspondance.

- **Zone de recherche des correspondants des points d'intérêt** – Dans tous les cas et sans exception, il est préférable de faire une mise en correspondance de points d'intérêt de l'image de référence vers le 8-voisinage des points d'intérêt de l'autre image. Cela augmente le nombre de correspondances trouvées et donc les chances d'obtenir une meilleure répartition des germes.
- **Seuil sur le score de corrélation** – Les seuils médians de chaque détecteur utilisé sont donnés dans la table 4.10. Ces résultats sont donnés à titre indicatif car les intervalles

DÉTECTEUR	SEUIL MÉDIAN
CENSUS	0.763 ± 0.037
GC	0.505 ± 0.085
SAD	0.881 ± 0.015
SMPD ₂	0.738 ± 0.048
NCC	0.98 ± 0.007

TABLE 4.10: *Seuils médians des mesures de corrélation.*

calculés pour recadrer entre 0 et 1 les seuils des mesures non normalisées ont été définis de manière empirique, voir § 4.2.1.3. Cependant, on peut remarquer que le seuillage est réalisé avec des seuils qui se situent au delà de 0.5, c'est-à-dire des valeurs plutôt stricts.

4.2.2.3 Reclassement des couples détecteur-corrélation selon le rang de compromis A après calcul des germes avec les paramètres médians

Les résultats de la méthode d'optimisation nous donnent des indications sur le paramétrage des couples détecteur-corrélation. Nous avons fixé chaque paramètre en prenant la valeur médiane des valeurs estimées pour chaque couple d'images. En conséquence, les résultats obtenus avec ces paramètres diffèrent des résultats précédents. Nous proposons alors de réévaluer les nouveaux germes obtenus selon le même critère, c'est-à-dire le compromis A, mais en calculant les germes avec les mêmes paramètres pour toutes les images. Le résultat est présenté dans la table 4.11.

Les trois premiers couples détecteur-corrélation restent les mêmes, mais dans un ordre différent. La méthode FAST-GC est celle qui donne le plus souvent les meilleurs résultats. Néanmoins, la méthode FAST-CENSUS donne en moyenne un meilleur compromis A. De plus, on constate qu'on perd en moyenne 5.7% de compromis A par rapport aux compromis obtenus avec les paramètres déterminés indépendamment pour chaque image.

COUPLE	RANG	C_A (%)	T_{ra} (%)	$T_{ré}$ (%)
FA-GC	4.0±1.7	35.9±13.3	77.5±11.6	45.4±12.0
FA-CENSUS	4.5±5.5	41.6±16.6	85.6±8.3	48.0±16.3
SU-CENSUS	6.8±5.3	37.7±17.4	82.2±7.8	44.3±17.1
HA.2-CENSUS	7.8±4.7	31.9±8.8	83.1±7.7	37.8±7.5
MIC-CENSUS	10.0±4.5	29.8±7.5	78.5±9.1	37.4±5.6
HA.3-CENSUS	10.8±9.5	30.9±9.5	83.2±6.6	36.5±8.3
HA.1-CENSUS	12.5±7.2	30.5±11.2	78.4±9.8	37.9±9.3
KR-CENSUS	12.8±7.6	30.4±11.5	86.7±8.1	34.4±10.6
MO-CENSUS	16.0±9.8	25.4±4.9	85.6±8.0	29.5±4.2
HA.1-SAD	17.8±6.7	28.4±11.2	76.6±11.5	35.9±9.4

TABLE 4.11: *Reclassement des couples détecteur-corrélation par rang de compromis A pour les germes calculés avec les paramètres médians. – Ce tableau présente les valeurs moyennes, suivies de l'écart-type, sur l'ensemble des couples d'images testés pour les 10 premiers couples détecteur-corrélation.*

On constate sur la vue globale présentée sur la figure 4.7, que les couples détecteur-corrélation avec les détecteurs multi-échelle HARRIS-LAPLACE, HESSIAN-LAPLACE et SURF peuvent donner de bons taux de rappel. Toutefois, ils donnent les taux de répartition les moins élevés, ce qui induit un mauvais compromis. Les taux de répartition les plus élevés sont obtenus avec les détecteurs SUSAN et FAST. De plus, ils donnent également un taux de rappel élevé, ce qui montre une nouvelle fois qu'ils permettent d'obtenir le meilleur compromis. De plus, cette figure met en évidence que la mesure de corrélation CENSUS est globalement plus performante que les autres. En effet, on constate que la majorité des pastilles bleues, qui représentent les couples qui utilisent CENSUS, sont situées au dessus des autres.

On peut donc dire que les couples détecteur-corrélation :

- FAST-CENSUS;
- FAST-GC et
- SUSAN-CENSUS

se détachent des autres pour donner le meilleur compromis entre le taux de rappel et le taux de répartition. Le premier donne le meilleur compromis moyen (41.6%), mais le second est le plus souvent mieux classé (rang moyen égal à 4). De plus, la mesure GC est moins coûteuse en temps de calcul que la mesure CENSUS. Le couple SUSAN-CENSUS donne également de bons résultats.

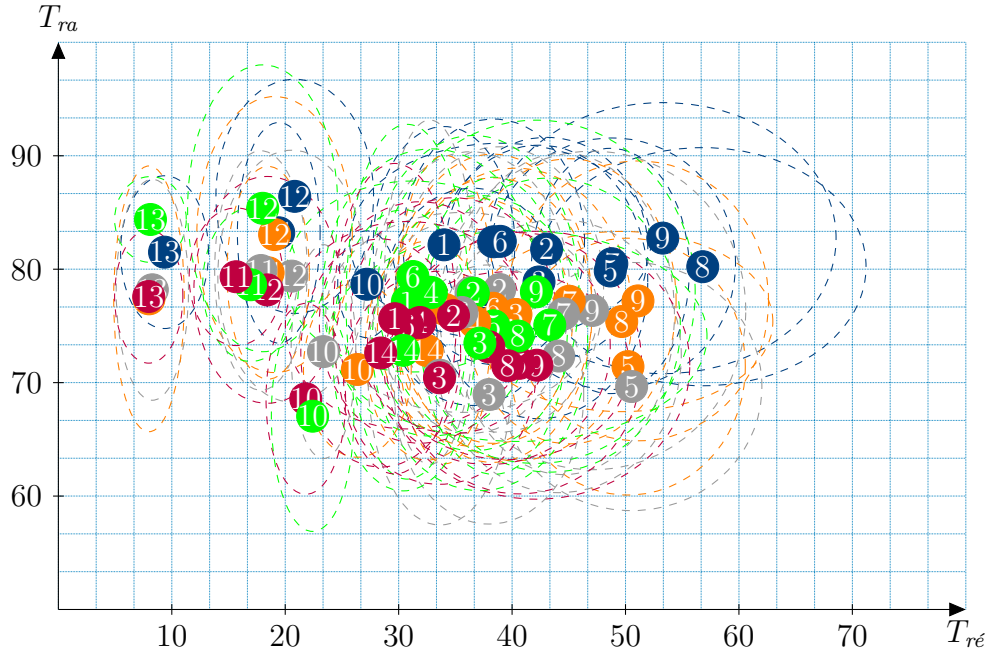


FIGURE 4.7: Vue d'ensemble des résultats pour le compromis A. – Chaque pastille correspond à un couple détecteur-corrélation. La position en abscisse correspond au taux de répartition moyen. La position en ordonnée correspond au taux de rappel moyen. Le numéro inscrit dans la pastille correspond au détecteur, voir § 4.2.1.2. Chaque couleur de pastille correspond à une mesure de corrélation (bleu = SAD, orange = GC, vert = NCC, violet = SMPD₂ et bleu = CENSUS). Les ellipses autour de chaque pastille indiquent l'écart-type. Là encore, on remarque bien que FAST-CENSUS (pastille bleue portant le numéro 9) obtient les performances les plus intéressantes.

4.3 Complémentarité des germes

Nous venons de déterminer que le meilleur compromis moyen entre le taux de rappel et le taux de répartition peut être obtenu en utilisant le couple détecteur-corrélation FAST-CENSUS et atteint 41.6% (avec un taux de rappel de 85.6% et une répartition de 48%). Nous considérons ce résultat comme solution « de référence ». L'objectif de cette section est de dépasser ce résultat. Il est possible de perfectionner le compromis A dans les cas suivants :

- Par amélioration du taux de répartition tel que :
 - ▶ le taux de rappel n'est pas dégradé ;
 - ▶ le taux de rappel est dégradé mais cette perte est compensée par l'amélioration du taux de répartition.

- Par amélioration du taux de rappel tel que :
 - ▶ le taux de répartition n'est pas dégradé ;
 - ▶ le taux de répartition est dégradé mais cette perte est compensée par l'amélioration du taux de rappel.

Pour cela, nous proposons de prendre l'union de différents ensembles de germes. Dans le pire des cas, le taux de répartition de l'union de deux ensembles vaut le plus petit taux parmi ceux de chacun des deux ensembles. Le taux de répartition ne peut donc être dégradé. Dans le meilleur des cas, on peut espérer obtenir pour l'union des deux ensembles un taux de répartition égal à la somme des taux de répartition des deux ensembles fusionnés.

En revanche, le taux de rappel de l'union de deux ensembles peut être dégradé. En effet, c'est par exemple le cas lorsque les deux ensembles ont en commun des germes corrects mais contiennent des germes erronés différents. Toutefois, si on fusionne deux ensembles qui ont des taux de rappel du même ordre et si ce taux est élevé, alors on peut espérer que cette dégradation soit faible. La dégradation maximale est obtenue lorsque les germes erronés des deux ensembles sont tous différents alors que tous les germes corrects sont identiques. Prenons par exemple deux ensembles de germes de même cardinalité constitués de 1000 germes, et ayant le même taux de rappel égal à 80%. Dans le pire des cas, seuls les germes erronés sont différents et répartis différemment dans les deux ensembles (c'est-à-dire dans des régions différentes). En prenant l'union des deux ensembles, on a donc 400 germes erronés et 800 germes corrects, d'où un taux de rappel de $800/(800 + 400) \approx 67\%$, d'où une perte sur le taux de rappel d'environ 13%. En outre, dans le cas inverse, où les germes corrects des deux ensembles fusionnés sont différents et répartis différemment alors que les germes erronés sont identiques, une amélioration du taux de rappel est possible. Sur notre exemple, l'amélioration du taux de rappel est de $1600/(1600 + 200) - 80\% \approx 9\%$. Néanmoins, on remarque que ces configurations sont des cas particuliers rarement rencontrés en pratique.

Ainsi, on propose d'améliorer le compromis A en tentant d'améliorer le taux de répartition de deux manière différentes. La table 4.18 résume les meilleurs résultats obtenus pour chaque piste envisagée :

- **Piste 1** – On prend toutes les paires de couples détecteur-corrélation. Le compromis A est ensuite évalué sur chacune de ces paires. La table 4.12, présente les dix meilleures paires. On constate que l'union des couples SUSAN-CENSUS et FAST-CENSUS permet une nette amélioration du compromis A de +10.4% par rapport à la solution de référence. Ce gain est dû à la complémentarité de la répartition des germes des deux ensembles. En effet, le gain moyen de répartition est de +13.1%. Le gain moyen du taux de rappel est de -2.1%. C'est donc une légère perte compensée par le gain important sur le taux de répartition. Les 9 autres combinaisons présentées montrent également un gain sur le compromis, mais moins important.
- **Piste 2** – Pour ne pas risquer de trop dégrader le taux de rappel, on propose de

COUPLE A	COUPLE B	RANG	$C_A(\%)$	$T_{ra}(\%)$	$T_{ré}(\%)$
SU-CENSUS	FA-CENSUS	3.0±3.0	52.0±17.1	83.5±8.0	61.1±15.2
KR-CENSUS	FA-CENSUS	17.5±13.3	48.0±16.6	85.9±8.1	55.1±15.6
HA.2-CENSUS	FA-CENSUS	21.0±16.3	49.1±15.0	84.6±8.1	57.3±13.6
SU-GC	FA-CENSUS	22.5±10.4	49.1±13.7	81.6±9.2	59.3±11.7
HA.3-SAD	FA-CENSUS	29.0±28.2	48.6±15.5	83.3±8.4	57.4±13.9
HA.3-CENSUS	FA-CENSUS	30.0±40.0	49.4±15.1	84.6±7.7	57.7±13.5
MIC-SAD	FA-CENSUS	32.0±20.0	46.3±17.4	82.1±8.8	55.6±17.0
BE-CENSUS	FA-CENSUS	36.0±44.5	47.3±16.6	85.6±8.2	54.6±15.6
HA.2-GC	FA-CENSUS	39.0±24.5	46.0±16.7	83.0±8.8	54.7±16.2
MIC-CENSUS	FA-CENSUS	40.5±28.9	48.2±14.2	82.7±8.6	57.6±12.3

TABLE 4.12: Classement par rang médian des meilleures paires de couples détecteur-corrélation (piste 1).

fusionner tous les ensembles de germes dont le taux de rappel est supérieur ou égal à 85.6%, qui est le taux de rappel moyen pour le couple détecteur-corrélation qui donne le meilleur compromis moyen. La table 4.13 présente les couples détecteur-corrélation triés par taux de rappel. On fusionne donc les ensembles de germes obtenus avec les 5 premiers couples détecteur-corrélation de cette table. On obtient un compromis moyen de 53.7%, soit un gain moyen de +12.1%. Le taux de répartition de 61.9% est légèrement supérieur à celui trouvé avec la meilleure solution de la piste 1, et donne un gain moyen de +13.9%. En revanche, il n'y a pas de perte sur le taux de rappel moyen qui vaut 85.7%, d'où un gain de +0.1%.

L'autre possibilité pour tenter d'améliorer le compromis A est d'améliorer le taux de rappel sans que le taux de répartition ne soit dégradé (ou que sa dégradation soit compensée). On propose de fusionner des ensembles de germes dont les taux de rappel sont très élevés et même supérieurs aux taux donnant le meilleur compromis A. Pour cela, nous commençons par déterminer les ensembles de germes qui donnent le meilleur taux de rappel, puis nous prenons l'union de ces ensembles. Le taux de répartition n'étant plus considéré pour le calcul de ces germes, les ensembles trouvés ont donc de très bons taux de rappel avec des taux de répartition généralement plus faibles que précédemment. On peut espérer que l'union de ces nouveaux ensembles compense cette perte pour obtenir une répartition au moins égale à celle du couple détecteur-corrélation qui donne le meilleur compromis A, c'est-à-dire 48.0%, mais cette fois avec un taux de rappel supérieur à 85.6%.

Nous commençons tout d'abord par déterminer les paramètres de chaque couple détecteur-corrélation qui donne le meilleur le taux de rappel. Le protocole utilisé est identique à celui du calcul des paramètres donnant le meilleur compromis A. Mais, cette fois, le compromis

COUPLE (%)	RANG (%)	C_A (%)	T_{ra} (%)	$T_{ré}$ (%)
KR-CENSUS	12.8±7.6	30.4±11.5	86.7±8.1	34.4±10.6
SF-NCC	67.0±0.7	6.5±2.6	85.7±4.8	7.5±2.7
MO-CENSUS	16.0±9.8	25.4±4.9	85.6±8.0	29.5±4.2
HeL-CENSUS	52.5±1.1	14.1±4.2	85.6±8.3	16.2±3.5
FA-CENSUS	4.5±5.5	41.6±16.6	85.6±8.3	48.0±16.3
BE-CENSUS	21.0±11.6	25.6±8.1	85.5±8.5	29.7±7.6
SF-SAD	65.8±3.3	6.6±3.0	85.0±6.0	7.6±3.0
HeL-NCC	53.0±10.4	13.7±6.0	83.8±7.8	15.9±5.6
HL-CENSUS	48.3±5.3	15.1±4.7	83.3±9.2	17.7±3.7
HA.3-CENSUS	10.8±9.5	30.9±9.5	83.2±6.6	36.5±8.3

TABLE 4.13: *Reclassement des couples détecteur-corrélation par compromis A où les germes sont calculés en prenant les paramètres médians optimaux (tri par taux de rappel). – Ce tableau présente les valeurs moyennes, suivies de l'écart-type, sur l'ensemble des couples d'images testés, des 10 premiers couples détecteur-corrélation selon le compromis A.*

recherché est le suivant :

$$C_B(\mathcal{G}) = T_{ra}(\mathcal{G}) \times \text{Coefficient de cardinalité}(\mathcal{G}) \quad (4.6)$$

Nous souhaitons obtenir le meilleur le taux de rappel mais nous pensons qu'il est important de le pondérer par un coefficient de cardinalité (d'autant plus grand que le nombre de germes corrects est élevé) pour éviter, par exemple, d'avoir des solutions avec un seul et unique germe correct. Le nombre maximal de germes obtenus dépend du paramétrage utilisé pour le calcul des germes, mais, bien sûr, aussi du couple d'images. Par exemple, avec un même paramétrage, s'il existe de nombreuses zones homogènes dans les images considérées, il est normal d'obtenir moins de points d'intérêt qu'avec un couple qui présente de nombreuses zones texturées. En pratique, nous avons pu constater, par exemple, qu'il est rare de dépasser 2000 germes pour l'image *Cones* de 375×450 pixels (soit 1%). Par conséquent, nous proposons que le coefficient de cardinalité reflète ce phénomène en choisissant une fonction logarithmique qui vaut 0 lorsque le nombre de germes corrects est nul et 1 lorsqu'il atteint le nombre de pixel de l'image :

$$\text{Coefficient de cardinalité}(\mathcal{G}) = \frac{\ln \left(\text{card} \left\{ \mathbf{g}_{i,j} \in \mathcal{G} \mid \text{Err} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \leq 0.5 \right\} + 1 \right)}{\ln(hw + 1)}. \quad (4.7)$$

Cette fonction est représentée sur la figure 4.8.

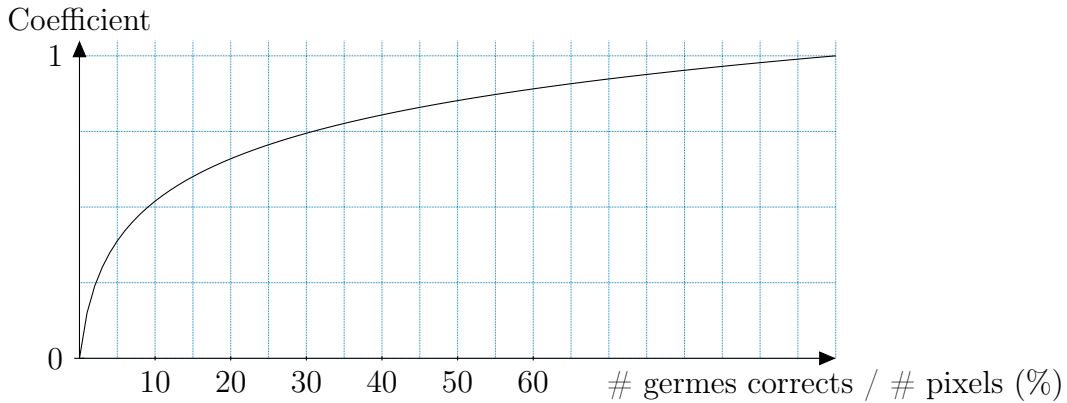


FIGURE 4.8: Coefficient de cardinalité des germes corrects. – Ce coût croît avec le nombre de germes corrects (illustré en % du nombre de pixels de l'image). Cette croissance est plus lente quand on s'approche de 100% des pixels de l'image.

Pour illustrer le compromis B, nous pouvons donner l'exemple suivant. Si nous trouvons 1000 germes corrects, donnant un coefficient de cardinalité de 80%, le taux de rappel permet de pondérer ce coefficient en fonction du rapport entre le nombre de germes corrects et le nombre de germes erronés. Si, à côté de ces 1000 germes corrects, il n'y a pas de germes erronés, alors le taux de rappel est égal à 100%, d'où un compromis B qui vaut 80%. En revanche, s'il existe en plus 600 germes erronés, on a un taux de rappel de $1000/(1000 + 600) = 62.5\%$, d'où un compromis B égal à 50%. Les paramètres médians pour les 10 meilleurs couples détecteur-corrélation selon le compromis B sont donnés dans la table 4.15. Il est intéressant de noter que le paramétrage du détecteur est, comme pour le compromis A, tel que le nombre de points d'intérêt est élevé, c'est-à-dire une faible échelle et un seuil quasi-nul. En revanche, le seuil sur la mesure de corrélation un petit peu plus strict que pour le compromis A. Enfin, la mise en correspondance est effectuée 6 fois sur 10 avec une zone de recherche réduite aux points d'intérêt (on ne considère pas leurs voisins). On peut donc penser que le fait de prendre en compte le voisinage des points d'intérêt rajoute des candidats et donc de l'ambiguïté et de l'imprécision.

Ainsi, on propose d'améliorer le compromis A en tentant d'améliorer le taux de rappel de deux manières différentes :

- **Piste 3** – On prend toutes les paires de couples détecteur-corrélation dont les paramètres ont été obtenus dans l'objectif d'obtenir le meilleur compromis B. Le compromis A est ensuite évalué sur chacune de ces paires. La table 4.16 présente les 10 meilleures unions. La meilleure union est obtenue avec les couples SUSAN-CENSUS et FAST-SAD si on considère le tri par rang. Si on regarde le compromis moyen, on constate que l'union de HARRIS3-CENSUS avec SUSAN-CENSUS donne un meilleur compromis A de 43.8%, soit une amélioration de +2.2%. Le taux de rappel moyen est de 87%, soit

une amélioration de +1.4%. La répartition est de 49.6%, soit un gain de +1.6%.

- **Piste 4** – Enfin, on propose de fusionner tous les ensembles de germes dont le taux de rappel est supérieur ou égal à 89.9%, qui est le taux de rappel moyen pour le couple détecteur-corrélation qui donne le meilleur compromis A moyen pour les germes calculés avec les paramètres médians pour le compromis B. La table 4.17 présente les couples détecteur-corrélation triés par taux de rappel. On fusionne les ensembles de germes obtenus avec les 9 premiers couples détecteur-corrélation de cette table. On obtient un compromis moyen de 51%, soit un gain de +9.4%. Le taux de rappel moyen est de 91%, soit un gain de +5.4%. Enfin le taux de répartition moyen est de 56%, soit un gain de +8%. C'est donc la solution qui permet d'obtenir un bon compromis entre le taux de rappel et le taux de répartition tout en assurant un taux de rappel très élevé.

COUPLE (%)	RANG (%)	C_B (%)	C_A (%)	T_{ra} (%)	$T_{ré}$ (%)
FA-CENSUS	1.8±0.8	50.0±6.7	40.4±17.6	89.9±6.7	43.9±16.9
MIC-CENSUS	4.5±2.9	48.3±7.9	35.4±19.8	92.5±4.9	37.4±19.6
KR-CENSUS	4.8±4.3	49.0±5.5	37.9±17.9	89.3±4.2	41.8±18.5
SU-CENSUS	5.5±5.6	48.9±8.6	38.4±19.5	92.3±3.7	40.9±20.1
BE-CENSUS	6.0±2.5	47.4±5.9	31.0±12.7	91.2±5.2	33.3±12.3
FA-SAD	8.0±1.7	46.6±6.7	34.5±19.0	88.4±4.9	37.9±19.4
SU-GC	8.8±5.9	46.5±6.3	31.0±11.4	87.1±6.6	34.8±11.2
FA-GC	11.0±9.9	46.3±8.6	34.9±13.9	85.6±12.4	39.7±12.3
KR-SAD	11.3±1.9	45.5±6.8	27.6±14.1	88.8±3.9	30.5±14.7
KR-GC	11.5±4.0	45.6±7.3	31.0±13.3	85.5±8.2	35.3±13.0

TABLE 4.14: Classement des couples détecteur-corrélation par rang de compromis B. – Ce tableau présente les valeurs moyennes, suivies de l'écart-type, sur l'ensemble des couples d'images, des 10 premiers couples détecteur-corrélation selon le rang de compromis B.

4.4 Conclusion

Nous avons proposé dans ce chapitre une étude de la sélection de germes par appariement de points d'intérêt pour la mise en correspondance par propagation en stéréovision

COUPLE	ÉCHELLE	SEUIL	FENETRE	SEUIL	ZDR
KR-SAD	$n \times n=5 \times 5$	0	11×11	0.895	✓
KR-GC	$n \times n=5 \times 5$	0	11×11	0.67	✓
BE-CENSUS	$n \times n=5 \times 5$	0	11×11	0.792	
MIC-CENSUS	$n \times n=5 \times 5$	0.01	11×11	0.8	
SU-GC	$r=2$	0	11×11	0.625	
SU-CENSUS	$r=2$	0	11×11	0.8	✓
FA-SAD	$r=2$	0.02	11×11	0.885	
FA-GC	$r=2$	0.02	11×11	0.595	
FA-CENSUS	$r=2$	0.01	11×11	0.817	

TABLE 4.15: Paramètres des meilleurs couples détecteur-corrélation pour obtenir le meilleur compromis B.

COUPLE A	COUPLE B	RANG	C_A (%)	T_{ra} (%)	$T_{ré}$ (%)
SU-CENSUS	FA-SAD	12.0 ± 13.3	43.0 ± 14.0	85.6 ± 7.6	49.2 ± 12.8
HA.3-CENSUS	SU-CENSUS	21.0 ± 25.2	43.8 ± 11.1	87.0 ± 6.9	49.6 ± 9.0
HA.3-CENSUS	FA-SAD	36.5 ± 48.9	43.5 ± 13.6	84.3 ± 7.3	50.8 ± 12.6
SU-CENSUS	FA-CENSUS	37.5 ± 5.9	36.5 ± 8.8	89.8 ± 7.9	40.6 ± 8.5
HA.3-CENSUS	FA-CENSUS	45.0 ± 34.1	41.6 ± 8.7	87.9 ± 6.6	47.0 ± 7.2
SU-GC	FA-SAD	59.0 ± 62.3	39.9 ± 12.9	84.3 ± 7.9	46.3 ± 11.8
MO-SAD	FA-SAD	64.5 ± 67.5	39.5 ± 14.2	83.4 ± 9.4	46.2 ± 13.1
FA-SAD	MO-SAD	64.5 ± 67.5	39.5 ± 14.2	83.4 ± 9.4	46.2 ± 13.1
HA.3-CENSUS	BE-CENSUS	68.0 ± 26.7	40.2 ± 9.1	86.9 ± 7.1	45.8 ± 7.3
HA.2-SAD	FA-SAD	71.0 ± 68.9	39.6 ± 13.3	82.8 ± 8.4	46.8 ± 12.4

TABLE 4.16: Classement des meilleures paires de couples détecteur-corrélation (piste 3).

COUPLE	C_A (%)	T_{ra} (%)	$T_{ré}$ (%)
MIC-CENSUS	15.3±7.1	92.9±8.7	17.2±9.5
FA-CENSUS	25.6±6.2	92.1±7.9	28.1±7.6
HeL-SMPD ₂	11.0±4.7	92.1±7.0	11.7±4.2
HA.2-CENSUS	25.1±7.7	91.9±5.7	27.2±7.8
KR-CENSUS	22.3±8.8	91.6±7.2	24.6±10.2
HL-SMPD ₂	8.1±2.7	90.8±6.2	8.9±2.9
MIC-SMPD ₂	17.0±8.4	90.7±11.2	18.8±8.4
HA.1-CENSUS	24.3±5.2	90.4±7.8	26.8±4.4
MO-CENSUS	18.5±3.2	90.2±7.3	20.5±2.9
BE-CENSUS	22.7±10.9	89.7±8.1	25.3±11.2

TABLE 4.17: *Reclassement des couples détecteur-corrélation par compromis B pour les germes calculés avec les paramètres médians optimaux (tri par taux de rappel). – Ce tableau présente les valeurs moyennes, suivies de l'écart-type, sur l'ensemble des couples d'images testés, des 10 premiers couples détecteur-corrélation selon le compromis A.*

binoculaire. Il existe de nombreuses méthodes de détection de points d'intérêt dont le comportement dépend du paramétrage effectué. Une difficulté réside dans ce paramétrage afin d'obtenir les meilleurs résultats selon un critère défini. Pour la mise en correspondance par propagation de germes, nous avons besoin d'appariements fiables et bien répartis. Nous avons défini un compromis qui est élevé lorsque ces deux critères sont élevés. Nous avons déterminé les paramètres des différentes méthodes d'appariement de points d'intérêt qui permettent de satisfaire au mieux ce compromis. Nous avons ensuite évalué et comparé les différents ensembles de germes obtenus avec les meilleures méthodes et nous avons proposé de combiner les ensembles les plus complémentaires pour obtenir de nouveaux ensembles donnant un compromis plus élevé. Les meilleurs ensembles obtenus sont présentés dans la figure 4.9 et les résultats sont résumés dans la table 4.18. Dans le chapitre suivant, nous réalisons des mises en correspondance par propagation à partir notamment de ces ensembles.

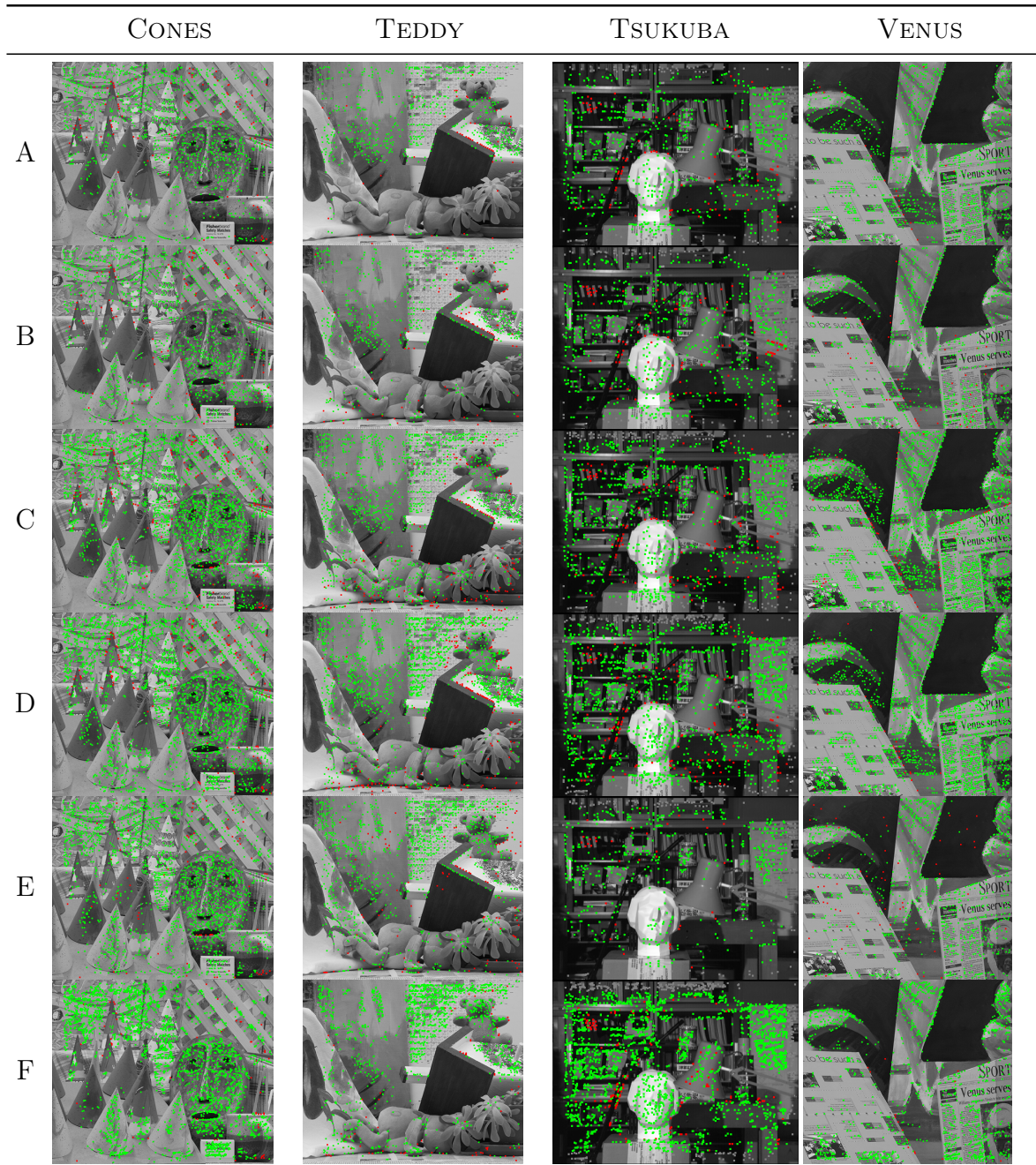


FIGURE 4.9: *Pixels gauches des germes des 6 meilleurs ensembles.* – Les points verts représentent les germes corrects, les points rouges les germes qui ne sont pas corrects et les points gris représentent les germes dont la disparité théorique n'est pas connue.

ABR.	C_A (%)	T_{ra} (%)	$T_{ré}$ (%)
Réf.	41.6±16.6	85.6±8.3	48±16.3
Piste 1	52.0±17.1 (+10.4)	83.5±8 (-2.1)	61.1±15.2 (+13.1)
Piste 2	53.7±17 (+12.1)	85.7±9.2 (+0.1)	61.9±14.3 (+13.9)
Piste 3	43±14 (+2.2)	85.6±7.6 (+1.4)	49.2±12.8 (+1.6)
Piste 4	51±12.3 (+9.4)	91±8.6 (+5.4)	56±11.3 (+8)

TABLE 4.18: *Performances des ensembles de germes complémentaires.* – Ce tableau reprend les taux de rappel et de répartition obtenus pour les meilleurs résultats de chaque piste de complémentarité envisagée. Le gain par rapport à la solution de référence est noté entre parenthèses. Les termes en gras correspondent aux meilleurs résultats pour chaque critère.

5 Mise en correspondance par propagation multi-mesure de germes

LES méthodes de mise en correspondance par propagation de germes donnent généralement un résultat plus ou moins dense selon les valeurs de différents paramètres. L'objectif de la mise en correspondance par propagation multi-mesure de germes est de calculer un ensemble d'appariements qui constitue un meilleur compromis entre le taux d'appariements corrects et la densité que lorsqu'on utilise une seule mesure. Les travaux décrits dans [Chambon 11] ont montré que les mesures de corrélation robustes aux discontinuités de profondeur donnent de bons résultats dans ces zones, mais également que, dans des zones sans difficulté, elles sont moins performantes que les mesures de corrélation classiques. Par conséquent, dans le but de combiner les avantages de différentes mesures de corrélation, nous proposons dans ce chapitre une technique de mise en correspondance par propagation multi-mesure. L'idée est de commencer par propager les germes, qui *a priori* ne sont pas situés dans les zones de discontinuité de profondeur, avec une mesure de corrélation classique et d'utiliser, sur le score de corrélation, une contrainte de seuil stricte pour arrêter la propagation dès que les voisinages des correspondants diffèrent trop. On suppose alors que les pixels restant à apparier se situent dans les zones proches des discontinuités de profondeur et on utilise une mesure de corrélation robuste pour poursuivre la propagation. Le principe est illustré par la figure 5.1 où les disparités en vert sont obtenues avec une mesure classique et celles en bleu avec une mesure robuste. Nous remarquons que ces dernières concernent presque partout les pixels situés dans les zones proches des discontinuités de profondeur. En adaptant la mesure de corrélation au contexte, on souhaite obtenir des résultats plus denses qu'avec une seule mesure sans faire décroître le pourcentage d'appariements corrects. De plus, la technique multi-mesure pro-

posée peut s'adapter aussi bien à l'approche simultanée qu'à l'approche séquentielle de la mise en correspondance par propagation de germes présentées § 5.3.1.2, page 139.

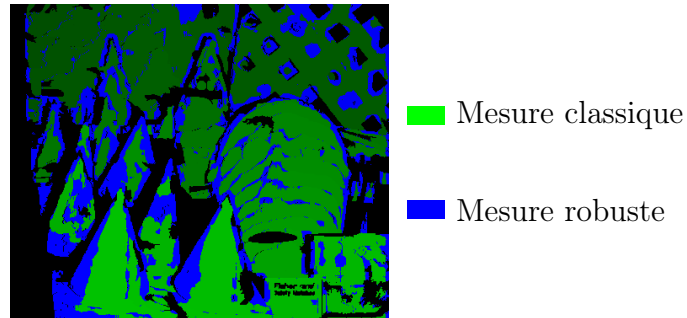


FIGURE 5.1: *Exemple de disparités obtenues par une mesure classique d'abord puis par une mesure robuste lors d'une propagation multi-mesure sur le couple Cones.*

5.1 Algorithme de propagation multi-mesure de germes

Dans cette partie, nous détaillons l'approche multi-mesure proposée après avoir présenté les méthodes de sélection de germes que nous avons choisies.

5.1.1 Sélection des germes

Au chapitre précédent, nous avons évalué différentes méthodes de sélection des germes par appariements de points d'intérêt. La répartition la plus élevée obtenue atteint 61.9%. Dans les cas où la densité du résultat final est la priorité, nous pensons que ce score peut ne pas être suffisant. Par conséquent, nous proposons une étape de complètement de germes pour pallier ce manque et une étape de validation pour éliminer certaines erreurs. La sélection des germes est donc constituée des trois étapes suivantes :

1. **Sélection des germes par appariements de points d'intérêt** – Il s'agit de la technique détaillée au chapitre 3, page 75.

2. **Complètement par sélection d'appariements d'intérêt** – La méthode de sélection d'appariements d'intérêt est généralement une alternative à l'appariement de points d'intérêt, mais elle est utilisée ici pour compléter les germes obtenus par appariement de points d'intérêt. Ces derniers sont marqués d'un point noir cerclé sur la figure 5.2. Nous utilisons une segmentation fine en régions de couleurs homogènes de l'image, illustrée par des traits épais sur la figure 5.2. Pour chaque région où il n'y a pas de germe obtenu par appariement de points d'intérêt, nous effectuons une recherche exhaustive des appariements de tous les pixels de la région, en parcourant la zone de recherche constituée des pixels de l'image droite situés sur la même ligne que le pixel considéré, et nous conservons les meilleurs selon le score de corrélation. Cette étape permet d'améliorer la répartition car elle ne fait que rajouter des germes. Ces germes sont marqués d'un point bleu sur la même figure. Cependant, de nouvelles erreurs risquent d'être introduites et d'entraîner une baisse du taux de rappel. Pour limiter les risques d'obtenir de faux appariements, en particulier dans les zones proches des discontinuités de profondeur, nous pensons qu'il est préférable d'utiliser une mesure de corrélation robuste. La contrainte de seuil n'est pas appliquée puisque, dans les zones difficiles, il est possible que deux correspondants donnent un score de corrélation faible. En revanche, les appariements qui ne vérifient pas la contrainte de symétrie sont éliminés.
3. **Validation** – Cette étape permet ensuite d'éliminer les appariements les moins probables. Nous faisons l'hypothèse que les germes d'une même région appartiennent à une même surface et doivent donc avoir des disparités proches. Une segmentation, moins fine que celle utilisée pour l'étape précédente, est utilisée pour calculer la médiane et l'écart-type robuste des disparités des germes calculés. L'utilisation des estimateurs robustes de la tendance centrale et de la dispersion des disparités permet d'éviter l'influence des disparités aberrantes. Les germes dont la disparité est trop éloignée de la médiane sont éliminés. Cette étape est illustrée par la figure 5.2 où les germes éliminés sont barrés d'une croix rouge. Le fait d'utiliser une segmentation moins fine permet de prendre en compte davantage de germes dans le calcul de la médiane et de l'écart-type. Par ailleurs, les erreurs sont généralement concentrées aux mêmes endroits (dans les zones difficiles) et, si nous utilisons une segmentation trop fine, le risque de ne considérer que les erreurs est plus grand.

5.1.2 Propagation multi-mesure

La propagation se déroule ensuite en deux étapes successives :

1. **Propagation avec mesure classique** – Les germes sont propagés avec une mesure de corrélation classique. Ainsi nous mettons d'abord en correspondance les pixels les

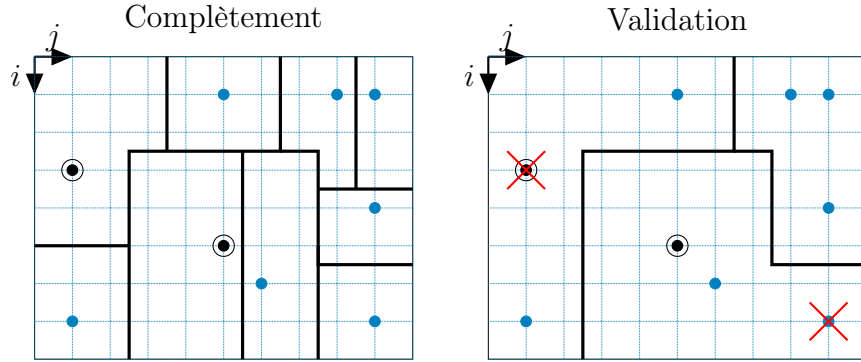


FIGURE 5.2: *Complètement et validation des germes.*

plus « faciles » à apparier. Un seuil strict sur le score de corrélation permet d'écarter les couples les moins ressemblants. Cela ne signifie pas forcément que ces couples ne se correspondent pas, mais que leur voisinages sont différents. Nous supposons alors qu'il s'agit des pixels situés dans les zones proches des discontinuités de profondeur. Dès qu'il n'est plus possible de trouver de nouveaux appariements qui satisfont cette contrainte, la première étape de propagation s'arrête et le dernier ensemble de germes est mémorisé.

2. **Propagation avec mesure robuste** – Les germes issus de l'étape précédente sont maintenant propagés avec une mesure de corrélation robuste dont le principe a été présenté § 2.2.3.1, page 43.

5.2 Critères d'évaluation

Les critères suivants sont utilisés pour évaluer les résultats des propagations obtenues avec la méthode proposée :

- **Densité D** – On évalue la densité du résultat en calculant le rapport du nombre de pixels mis en correspondance sur le nombre total de pixels de l'image :

$$D(\mathcal{M}) = \frac{\text{card } \mathcal{M}}{hw} \quad (5.1)$$

où \mathcal{M} est l'ensemble des appariements calculés.

- **Densité dans les zones difficiles D_{ZD}** – Nous évaluons également la densité du résultat dans les zones difficiles, c'est-à-dire les pixels proches des discontinuités de

profondeur et les pixels proches des zones occultées.

$$D_{\text{ZD}}(\mathcal{M}_{\text{ZD}}) = \frac{\text{card } \mathcal{M}_{\text{ZD}}}{\text{card } \text{ZD}} \quad (5.2)$$

où ZD désigne l'ensemble des pixels situés dans les zones difficiles, \mathcal{M}_{ZD} désigne l'ensemble des appariements obtenus dans ces zones. Les zones difficiles sont déterminées à partir de la vérité terrain en considérant qu'un pixel appartient à une zone difficile s'il n'est pas occulté et si, dans son voisinage, il existe une discontinuité de profondeur ou un pixel occulté. Les dimensions de ce voisinage sont données par les dimensions de la fenêtre de corrélation utilisée. La figure 5.3 montre en blanc les zones difficiles de l'image *Cones* pour une fenêtre de taille 11×11 .



FIGURE 5.3: Carte des zones difficiles (en blanc) du couple *Cones* pour une taille de fenêtre de corrélation 11×11 .

- **Taux de rappel** T_{ra} – Le taux de rappel est le même que pour les germes, c'est-à-dire le nombre d'appariements corrects sur le nombre d'appariements effectués :

$$T_{ra}(\mathcal{M}) = \frac{\text{card} \left\{ \mathbf{d}_{i,j} \in \mathcal{M} \mid \text{Err} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \leq 0.5 \right\}}{\text{card } \mathcal{M}}. \quad (5.3)$$

- **Taux d'acceptation** T_a – Cependant, nous pensons que l'évaluation des propagations peut être moins stricte que celle des germes. Nous utilisons le taux d'acceptation qui ressemble au taux de rappel, mais dont la tolérance aux erreurs est plus grande :

$$T_a(\mathcal{M}) = \frac{\text{card} \left\{ \mathbf{d}_{i,j} \in \mathcal{M} \mid \text{Err} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) \leq 1 \right\}}{\text{card } \mathcal{M}}. \quad (5.4)$$

- **Compromis C** C_C – Contrairement au cas des germes où on souhaitait obtenir à la fois un taux de rappel élevé et une répartition élevée, pour avoir le plus de germes fiables possibles, ici le compromis sur les propagations est un petit peu différent. En effet, selon l'application souhaitée, on peut avoir besoin de l'un ou de l'autre. On propose alors d'utiliser la moyenne des deux critères souhaités pour quantifier ce compromis :

$$C_C(\mathcal{M}, \mathcal{R}) = \frac{T_a(\mathcal{M}) + D(\mathcal{M})}{2}. \quad (5.5)$$

5.3 Évaluation de la propagation multi-mesure

Dans cette partie, nous évaluons la technique de propagation multi-mesure appliquée aux deux approches (séquentielle et simultanée). Ces propagations sont réalisées à partir des meilleurs ensembles de germes déterminés au chapitre 4, d'une part, et complétés et validés par la méthode proposée § 5.1.1, d'autre part.

5.3.1 Protocole expérimental

5.3.1.1 Images testées

Les images testées sont les mêmes que dans le chapitre précédent, voir § 4.2.1.1, page 108.

5.3.1.2 Paramètres fixés

Pour déterminer les paramètres qui permettent d'obtenir des résultats satisfaisant les compromis souhaités, une approche par optimisation analogue à celle proposée au chapitre 4 serait intéressante. Malheureusement, une telle approche est très coûteuse en temps de calculs. Nous avons préféré une étude plus traditionnelle en faisant varier indépendamment différents paramètres. Par ailleurs, on retrouve dans l'étude sur les couples détecteur-corrélation certains paramètres, en particulier ceux sur les mesures de corrélation. On décide donc de fixer ces paramètres en fonction des résultats précédemment obtenus. De plus, d'autres paramètres peuvent être fixés de façon à se concentrer sur ceux qui nous semblent jouer un rôle plus déterminant dans la propagation.

Les paramètres de propagation suivants sont fixés :

- **Fenêtre de voisinage** – Il s’agit des voisins autour des pixels d’un germe pour lesquels on recherche des correspondances. Nous considérons le 8-voisinage des germes.
- **Zone de recherche** – Comme les images testées ont subi une rectification épipolaire, nous pouvons en fait limiter la recherche des correspondants à deux candidats, voir la zone grisée sur la figure 5.8, page 148.
- **Mesure de corrélation classique** – Il faut choisir deux mesures de corrélation : une mesure « classique » et une mesure robuste. Par mesure classique, nous entendons ici une mesure performante dans les zones qui ne présentent pas de difficultés. Nous proposons de fixer la mesure classique mais de tester deux mesures robustes. Nous choisissons la mesure GC présentée § 2.2.3.1, page 43, qui a donné de bons résultats pour l’appariement des points d’intérêt (voir chapitre 4). La taille de la fenêtre de corrélation est fixée à 9×9 . Ces dimensions sont celles qui donnent les meilleurs résultats avec cette mesure sur les images testées.
- **Vérification bidirectionnelle souple** – Une vérification bidirectionnelle souple est employée pour éliminer les appariements qui ne vérifient pas la contrainte de symétrie.

5.3.1.3 Paramètres recherchés

Les paramètres de propagation suivants sont évalués :

- **Approches de propagation** – Les deux approches de propagation détaillées § 2.2.4, page 51, sont étudiées. Il s’agit de :
 - ▶ *l’approche séquentielle* ;
 - ▶ *l’approche simultanée*.
- **Ensemble initial de germes** – Nous testons les meilleurs ensembles de germes trouvés lors de l’évaluation présentée au chapitre 4. Ils sont résumés dans la table 5.1. Nous com-

ABR. GERMES	
A	FAST-CENSUS
B	FAST-GC
C	FAST-CENSUS \cup SUSAN-CENSUS
D	KR-CENSUS \cup SURF-NCC \cup MORAVEC-CENSUS \cup HESSIAN-LAPLACE-CENSUS \cup FAST-CENSUS
E	SUSAN-CENSUS \cup FAST-SAD
F	MIC-CENSUS \cup FAST-CENSUS \cup HESSIAN-LAPLACE-SMPD ₂ \cup HARRIS2-CENSUS \cup KITCHEN-CENSUS \cup HARRIS-LAPLACE-SMPD ₂ \cup MIC-SMPD ₂ \cup HARRIS1-CENSUS \cup MORAVEC-CENSUS

TABLE 5.1: *Ensembles de germes testés.*

plétons ces ensembles avec la méthode proposée § 5.1.1 en testant les mesures SMPD₂ et CENSUS pour le calcul des appariements d'intérêt. La segmentation fine utilisée pour la recherche de ces appariements est calculée par *mean-shift* avec les paramètres $h_s = h_r = 1$. La segmentation moyenne utilisée pour la validation est obtenue avec les paramètres $h_s = h_r = 8$. Ces segmentations ont donné des résultats satisfaisants lors de tests préliminaires.

- **Mesure de corrélation robuste** – Nous choisissons de tester la mesure robuste SMPD₂. Il s'agit de la meilleure mesure de la famille robuste selon l'évaluation présentée dans [Chambon 11]. Nous choisissons également de tester la mesure CENSUS présentée § 2.2.3.1, page 43. Cette dernière a donné les meilleurs résultats lors de l'appariement des points d'intérêt (voir chapitre 4). En outre, elle donne également les meilleurs résultats dans les zones difficiles selon l'évaluation présentée dans [Chambon 11]. La taille de la fenêtre de corrélation pour ces mesures est fixée à 11×11 . Ces dimensions sont celles qui donnent les meilleurs résultats avec ces mesures sur les images testées dans l'étude de la sélection des germes.
- **Seuil sur le score de corrélation des mesures classique et robuste** – Le premier seuil est celui qui arrête l'utilisation de la mesure classique pour permettre à la mesure robuste d'affiner les résultats dans les zones difficiles. En effet, dès qu'il n'existe plus d'appariement tel que le score de corrélation est supérieur à ce seuil, la propagation avec la mesure classique s'arrête et la mesure robuste prend le relais. S'il est trop strict, très peu de correspondances sont calculées avec la mesure classique. Inversement, s'il est trop tolérant, le risque est de ne laisser aucun appariement à calculer à la mesure robuste. Les intervalles des seuils testés ont été déterminés à partir de tests préliminaires. Dans le but de faciliter l'interprétation des résultats, nous utilisons les valeurs pseudo-normalisées, voir § 4.2.1.3, page 110.
- **Segmentation pour la contrainte de région** – Nous avons utilisé la contrainte de région avec différentes segmentations. Le choix de chaque niveau de segmentation s'appuie sur un critère visuel :
 - ▶ *Pas de segmentation* S_\emptyset – Pour évaluer l'utilité de la contrainte de région, nous réalisons également des tests sans cette contrainte.
 - ▶ *Segmentation fine* S_1 – Il s'agit de la même segmentation utilisée pour le calcul du taux de répartition des germes corrects ($h_s = h_r = 4$). Plus la segmentation est fine, plus elle contraint la propagation et plus elle limite la densité car il est alors plus difficile d'obtenir des germes dans toutes les régions.
 - ▶ *Segmentation moyenne* S_2 – Nous proposons d'utiliser une segmentation moins fine ($h_s = h_r = 8$) que celle utilisée pour évaluer le taux de répartition.
 - ▶ *Segmentation grossière* S_3 – Une segmentation plus grossière ($h_s = 48, h_r = 24$) est également testée. Plus la segmentation est grossière, moins la propagation est contrainte (puisque les régions sont plus grandes) et plus on favorise la propagation (et donc l'augmentation de la densité), tout en risquant de faire davantage d'erreurs

(et de les propager).

Les différentes segmentations utilisées pour la propagation sont présentées dans la figure 5.4.

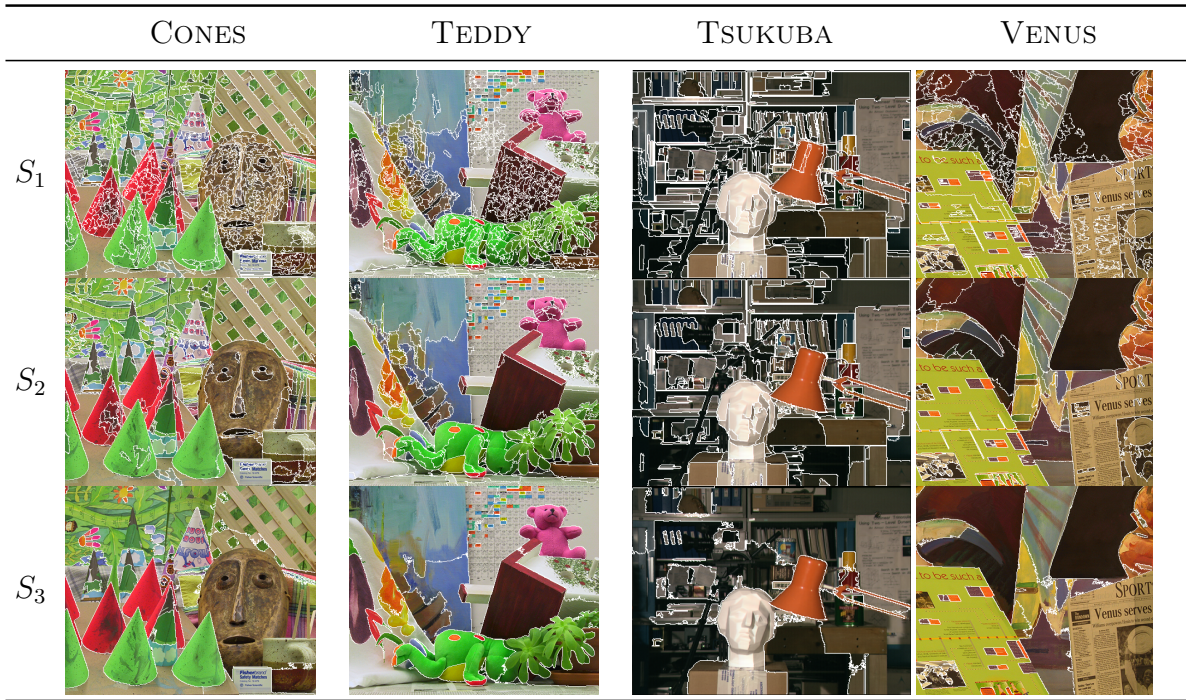


FIGURE 5.4: *Segmentations utilisées pour la contrainte de région lors des propagations.*

5.3.1.4 Résultats

Complètement et validation

La table 5.2 résume les différents taux de compromis A, voir définition (4.5), page 107, de rappel et de répartition des germes initiaux donnés au chapitre précédent ainsi que les taux obtenus après complètement par appariement d'intérêt et validation. On note $X + M$ un ensemble de germes X complétés par les appariements d'intérêt obtenus avec la mesure M et validés.

Nous constatons une amélioration du compromis A allant de +8.3% à +19%. Ce gain est obtenu grâce à l'augmentation de la répartition allant de +13.4% à +19.2%. Les

taux de rappel sont du même ordre avant et après le complètement avec parfois un léger gain. Nous remarquons également que les ensembles C et D qui avaient les meilleurs taux de répartition ont à nouveau les meilleurs taux de répartition après complètement et validation. De même, l'ensemble F obtient le meilleur taux de rappel avant et après complètement et validation.

ABR.	COMPROMIS A (%)	RAPPEL (%)	RÉPARTITION (%)
A	41.6±16.6	85.6±8.3	48±16.3
A+CENSUS	57.8±18.6	87.6±10.1	65.3±16.2
A+SMPD ₂	57.7±17.4	87.1±10	65.6±14.9
B	35.9±13.3	77.5±11.6	45.4±12.0
B+CENSUS	54.9±18.2	83.5±13.9	64.6±14.4
B+SMPD ₂	50±17.6	82±13.2	64.5±13.8
C	52.0±17.1	83.5±8.	61.1±15.2
C+CENSUS	62±18.4	85.7±9.2	71±15.7
C+SMPD ₂	61.4±18.6	84.9±9.3	71.3±15.3
D	53.7±17	85.7±9.2	61.9±14.3
D+CENSUS	62±17.1	86±9.1	71.2±13.7
D+SMPD ₂	62.3±16.5	90±9	71.6±12.6
E	43±14	85.6±7.6	49.2±12.8
E+CENSUS	57±17.2	86.9±8.8	65±14.7
E+SMPD ₂	60±17.4	86.6±9.2	65.7±14.6
F	51±12.3	91±8.6	56±11.3
F+CENSUS	62.4±13.7	90±8.3	68.7±10.8
F+SMPD ₂	63.2±13.7	91±8.5	69.4±10.4

TABLE 5.2: *Taux des ensembles de germes évalués.* – Ce tableau reprend les taux de rappel et de répartition obtenus pour les 6 ensembles de germes testés. Ces ensembles sont complétés par appariements d'intérêt (CENSUS et SMPD₂) et validation. Pour chaque critère, les termes en gras correspondent aux meilleurs résultats (avant et après complètement et validation).

Propagations de germes

La figure 5.5 présente les résultats des différents tests. Chaque point représente une mise en correspondance par propagation avec différents paramètres. La position de chaque point est donnée, en abscisse, par la densité du résultat et, en ordonnée par son taux d'acceptation. Les points violets correspondent aux résultats obtenus avec l'approche séquentielle et les points bleus à ceux obtenus avec l'approche simultanée. Nous avons également délimité par des lignes rouges (de façon manuelle) les propagations qui nous paraissent les plus intéressantes à étudier. Il s'agit de celles qui ont les taux d'acceptation ou les densités les plus élevés. En effet, selon le type d'application visé, il peut être intéressant d'avoir une forte densité avec le taux d'acceptation le plus élevé possible. Inversement, on peut souhaiter avoir un très fort taux d'acceptation quitte à avoir une densité plus faible.

Globalement, on constate sur la figure 5.5 que l'approche séquentielle permet d'obtenir des meilleurs résultats selon le taux d'acceptation mais avec des densités moins élevées qu'avec l'approche simultanée.

La figure 5.6 est identique à la figure 5.5 mais, cette fois, on montre en rouge les résultats obtenus sans la contrainte de région et en orange ceux obtenus avec une contrainte de région. On constate que, globalement, la contrainte de région permet d'obtenir des résultats avec un meilleur taux d'acceptation mais avec une densité moins élevée.

La figure 5.7 est identique à la figure 5.5 mais, cette fois, on montre en vert les résultats obtenus avec les mesures GC-CENSUS et en rose ceux obtenus avec les mesures GC-SMPD₂. On constate que les propagations réalisées par GC-CENSUS donnent globalement de meilleurs résultats que celles réalisées avec GC-SMPD₂.

La table 5.3 présente les résultats et les paramètres utilisés pour obtenir les solutions que nous considérons les plus intéressantes, c'est-à-dire au dessus de la limite indiquée en rouge sur les figures 5.5, 5.6 et 5.7. La colonne A indique l'approche utilisée (SI : simultanée, SE : séquentielle), la colonne G indique l'ensemble initial de germes utilisé, la colonne M indique la ou les mesure(s) utilisée(s) lors de la propagation (C. : CENSUS, S. SMPD₂), la colonne S indique le ou les seuil(s) utilisé(s) sur ces ou cette mesure(s) (\emptyset indique que la contrainte de seuil n'est pas utilisée) et la colonne R indique la segmentation utilisée pour la contrainte de région. Les propagations sont triées selon le compromis C.

Nous proposons de distinguer quatre classes qui sont présentées dans la table 5.3 :

- **C1** – Il s'agit des propagations qui donnent les densités les plus élevées, 90.6% en moyenne. Le taux d'acceptation moyen est égal à 91.1%. On constate que la densité dans les zones difficiles est relativement élevée, 87.87% en moyenne, mais le taux d'acceptation est relativement faible, 79.9% en moyenne. Cela montre que la plupart de erreurs se situent dans ces zones. Ces résultats sont obtenus avec l'approche simultanée à partir de l'ensemble de germes F (celui qui a le taux de rappel le plus élevé), complétés par SMPD₂ ou CENSUS, et validés. Seule la mesure CENSUS est utilisée lors de la propagation sans contrainte de seuil. La contrainte de région est utilisée avec la

segmentation la plus grossière (S_3) ou elle n'est pas utilisée (S_\emptyset).

- **C2** – Il s'agit des propagations qui donnent des densités élevées, 84.7% en moyenne, avec des taux d'acceptation élevés, 94.11% en moyenne. La densité dans les zones difficiles est moyennement élevée, 73.51% en moyenne, et le taux d'acceptation moyen y est égal à 84.97%. Le premier résultat de la classe C2 est obtenu avec l'approche simultanée à partir de l'ensemble de germes F, complété avec $SMPD_2$ et validé. La mesure CENSUS est utilisée seule, sans contrainte de seuil, mais avec une contrainte de région qui utilise une segmentation moyenne S_2 . Cela permet notamment d'éviter de propager certaines erreurs. Tous les autres résultats sont obtenus avec l'approche séquentielle à partir de tous les ensembles de germes (sauf le E), complétés par $SMPD_2$ et validés. La multi-mesure joue un rôle dans ces propagations. Un seuil relativement strict est utilisé sur la mesure GC et aucun seuil ou un seuil tolérant est utilisé sur CENSUS. Enfin la segmentation moyenne S_2 est utilisée par la contrainte de région.
- **C3** – Il s'agit des propagations qui donnent des densités moyennes, 64.8% en moyenne, avec des taux d'acceptation très élevés, 96.26% en moyenne. La densité dans les zones difficiles varie entre 39.3% et 66% avec une moyenne de 49.11%. Le taux d'acceptation moyen y est égal à 90.7%. Ces résultats sont tous obtenus avec l'approche séquentielle. Tous les ensembles de germes sont utilisés, sauf le B et le C. L'ensemble F est celui qui revient le plus souvent. Ces ensembles sont complétés par $SMPD_2$ sauf dans un cas où CENSUS est utilisée. La multi-mesure est utilisée 7 fois sur 17. On peut remarquer que plus le seuillage est strict sur CENSUS, plus la densité diminue. On remarque également que la contrainte de région est utilisée avec la segmentation la plus fine S_1 . Enfin, on remarque que lorsque CENSUS est utilisée seule, les résultats sont meilleurs dans les zones difficiles (53.09% en moyenne pour la densité et 91.49% en moyenne pour le taux d'acceptation des propagations qui utilisent uniquement CENSUS contre 43.43% en moyenne pour la densité et 89.57% en moyenne pour le taux d'acceptation des propagations multi-mesure).
- **C4** – Il s'agit des propagations qui donnent les taux d'acceptation les plus élevés, 97.9% en moyenne, mais aussi les densités les plus faibles, 47.4% en moyenne. La densité dans les zones difficiles est également faible, 27.28% en moyenne, mais le taux d'acceptation dans ces zones y est le plus élevé, 93.7% en moyenne. Ces résultats sont obtenus avec l'approche séquentielle à partir des ensembles de germes B, D et E, complétés par $SMPD_2$ ou CENSUS. La mesure CENSUS est utilisée seule et avec un seuil relativement strict. La contrainte de région n'est pas utilisée (sauf dans le premier cas où elle est utilisée avec la segmentation la plus fine S_1). On pense que la contrainte de seuil stricte arrête les propagations avant que les disparités propagées ne rencontrent les contours des régions, ce qui rend la contrainte de région inutile.

La figure 5.8 montre les cartes de disparité obtenues pour les premières propagations de chaque classe et pour la première propagation multi-mesure (de la classe C2). Les disparités erronées sont en rouge.

	A	G	M	S	R	C_C (%)	D (%)	T_a (%)	D_{ZD} (%)	T_{aZD} (%)
C1	SI	F+S.	C.	\emptyset	S_3	91.0±3.4	90.0±3.4	92.0±4.1	87.2±1.5	83.2±5.4
	SI	F+S.	C.	\emptyset	S_\emptyset	90.9±3.4	91.2±3.5	90.6±4.1	88.9±1.0	78.7±4.7
	SI	F+C.	C.	\emptyset	S_\emptyset	90.8±3.4	91.2±3.4	90.3±4.0	88.5±1.4	76.6±4.6
	SI	F+C.	C.	\emptyset	S_3	90.8±3.4	90.0±3.4	91.5±4.1	86.4±1.1	81.1±4.4
C2	SI	F+S.	C.	\emptyset	S_2	90.6±3.4	87.7±4.2	93.4±3.1	83.4±3.0	86.2±4.3
	SE	F+S.	GC-C.	(0.75, \emptyset)	S_2	90.0±3.4	86.0±4.8	94.1±2.2	73.9±3.2	85.4±3.8
	SE	D+S.	GC-C.	(0.75, \emptyset)	S_2	90.0±3.8	86.1±5.0	93.9±2.8	74.3±2.7	83.6±3.0
	SE	C+S.	GC-C.	(0.75, \emptyset)	S_2	89.9±4.0	85.9±5.4	93.8±2.9	74.1±3.2	84.0±3.9
	SE	A+S.	GC-C.	(0.75, \emptyset)	S_2	89.5±4.1	85.1±5.3	94.0±3.1	73.6±3.5	85.3±4.7
	SE	D+S.	GC-C.	(0.75, 0.65)	S_2	89.1±3.4	83.9±4.4	94.2±2.7	71.2±2.5	84.1±3.2
	SE	B+S.	GC-C.	(0.75, \emptyset)	S_2	89.0±4.4	83.8±6.4	94.2±2.6	73.1±4.4	85.0±2.9
	SE	C+S.	GC-C.	(0.75, 0.65)	S_2	88.9±3.7	83.4±5.0	94.3±2.6	71.0±3.2	84.4±4.0
	SE	F+S.	GC-C.	(0.75, 0.65)	S_2	88.7±2.5	82.6±2.7	94.7±2.5	70.3±3.2	85.9±4.0
	SE	A+S.	GC-C.	(0.75, 0.65)	S_2	88.5±3.8	82.5±5.0	94.5±2.9	70.2±2.7	85.8±4.6
C3	SE	F+S.	C.	\emptyset	S_1	85.8±5.2	75.9±8.7	95.7±2.3	66.0±7.8	90.0±3.1
	SE	F+S.	C.	0.65	S_1	85.0±5.1	74.0±8.5	95.9±2.2	63.8±7.4	90.4±3.3
	SE	F+S.	C.	0.7	S_1	83.5±4.5	70.7±7.6	96.2±2.1	59.8±7.0	91.0±3.3
	SE	D+S.	C.	0.75	S_1	82.2±2.9	68.2±4.8	96.2±2.0	57.2±5.3	90.5±3.0
	SE	F+S.	GC-C.	(0.75, 0.75)	S_1	81.8±3.6	67.3±6.4	96.4±1.9	49.0±5.2	89.4±3.6
	SE	F+S.	GC-C.	(0.5, 0.8)	S_1	81.8±4.7	67.1±8.6	96.4±1.6	43.1±5.8	87.3±4.4
	SE	F+S.	C.	0.75	S_1	81.5±4.1	66.3±6.9	96.6±2.1	53.5±6.6	91.8±3.5
	SE	A+S.	GC-C.	(0.75, 0.75)	S_1	80.5±3.8	64.6±6.4	96.5±2.0	48.6±5.8	89.9±4.6
	SE	A+S.	C.	0.75	S_1	80.2±3.5	63.9±5.9	96.5±1.9	52.5±7.1	91.3±3.9
	SE	D+S.	GC-C.	(0.75, 0.8)	S_1	79.7±2.7	62.8±4.3	96.6±1.8	43.2±2.5	89.0±3.3
	SE	F+S.	GC-C.	(0.75, 0.8)	S_1	79.2±3.4	61.6±6.1	96.8±1.7	40.7±4.7	90.0±3.4
	SE	D+C.	C.	0.8	S_1	79.1±2.4	61.6±4.4	96.7±1.8	45.7±4.2	91.4±4.3
	SE	D+S.	C.	0.8	S_1	79.0±2.5	60.8±4.1	97.1±1.8	45.6±4.5	92.7±3.7
	SE	E+S.	GC-C.	(0.75, 0.8)	S_1	78.6±3.0	60.5±4.7	96.8±2.0	39.3±3.8	90.8±4.7
	SE	F+S.	C.	0.8	S_1	78.5±3.7	59.8±6.6	97.2±1.9	44.1±6.9	92.8±4.0
	SE	A+S.	GC-C.	(0.75, 0.8)	S_1	78.0±3.4	59.1±5.3	97.0±1.8	40.1±4.3	90.6±4.7
	SE	A+S.	C.	0.8	S_1	77.4±2.7	57.6±4.3	97.2±1.8	42.7±6.0	93.0±4.7
C4	SE	D+S.	C.	0.85	S_1	73.6±2.5	49.6±4.1	97.7±1.8	31.7±4.5	94.1±4.3
	SE	B+C.	C.	0.85	S_\emptyset	72.6±2.6	47.2±4.6	97.9±1.4	26.2±4.0	93.2±5.0
	SE	E+C.	C.	0.85	S_\emptyset	72.2±2.4	46.5±4.4	98.0±1.2	25.7±3.9	93.7±4.1
	SE	E+S.	C.	0.85	S_\emptyset	72.2±2.4	46.3±4.4	98.0±1.3	25.5±3.9	93.8±4.4

TABLE 5.3: Classement par compromis C des meilleures propagations. – On indique en gras les meilleurs résultats pour chaque critère et pour chaque classe.

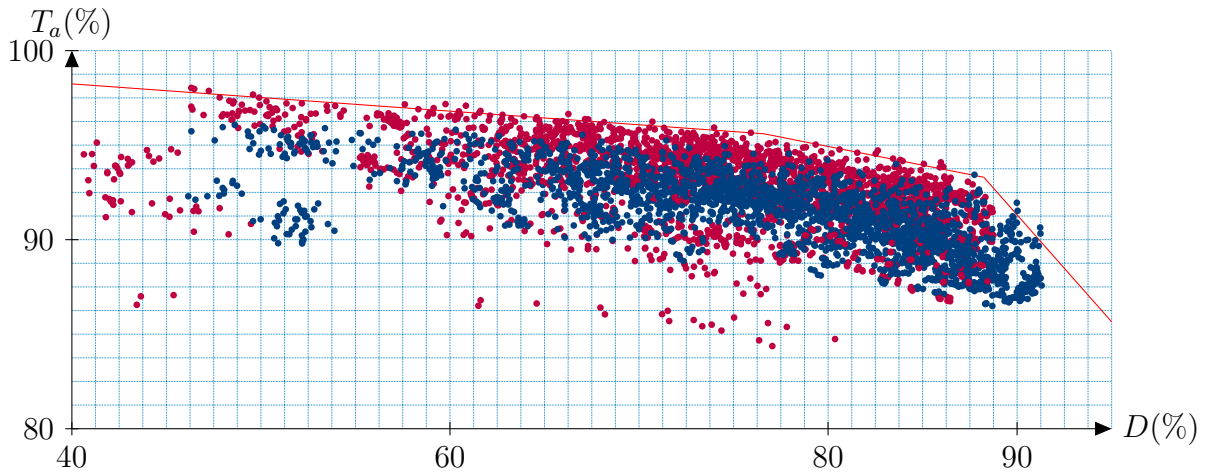


FIGURE 5.5: Taux d'acceptation en fonction de la densité des différentes propagations testées. – Les points bleus représentent les résultats de l'approche simultanée et les points violets ceux de l'approche séquentielle.

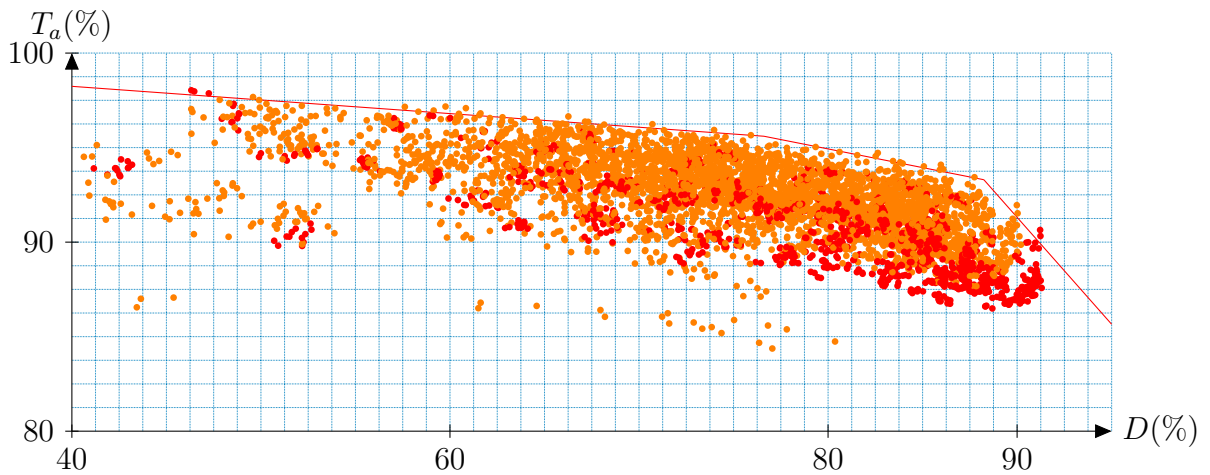


FIGURE 5.6: Taux d'acceptation en fonction de la densité des différentes propagations testées. – Les points rouges représentent les résultats obtenus sans la contrainte de région et les points orange ceux obtenus avec une contrainte de région.

5.3.2 Comparaison avec la mise en correspondance locale classique

Nous comparons les résultats obtenus par mise en correspondance par propagation avec ceux obtenus en utilisant l'approche classique *uta*, voir § 2.2.1, page 33. La figure 5.9 montre les résultats de la mise en correspondance *uta* avec la mesure CENSUS. Les

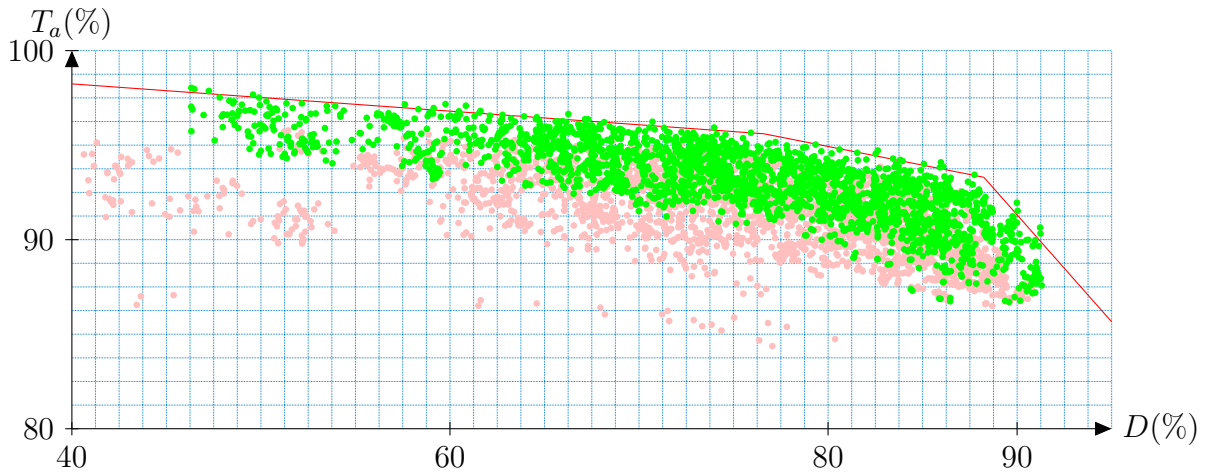


FIGURE 5.7: Taux d'acceptation en fonction de la densité des différentes propagations testées. – Les points verts représentent les résultats obtenus avec les mesures GC-CENSUS et ceux en rose ceux obtenus avec les mesures GC-SMPD₂.

disparités erronées sont indiquées en rouge. La table 5.4 présente les scores obtenus avec la méthode *wta* utilisant la mesure CENSUS, c'est-à-dire celle qui donne les meilleurs résultats. La contrainte de seuil est également utilisée, sa valeur est indiquée dans la colonne S (les mêmes valeurs de seuil sont utilisées pour la propagation). Sans contrainte de seuil, la densité obtenue est très proche de la propagation qui donne le compromis C le plus élevé (+0.07%) et le taux d'acceptation le plus élevé (+0.84%). La densité dans les zones difficiles est plus élevée (+1.8%) avec l'approche *wta*, mais le taux d'acceptation est plus faible (-0.68%). En utilisant la contrainte de seuil avec la méthode *wta*, on retrouve des compromis proches de certains obtenus par propagation. Cependant, si on se concentre sur les zones difficiles, on constate quelques différences. Avec l'approche *wta*, et un seuillage à 0.75, le taux d'acceptation dans les zones difficiles est de 86.42% et la densité y est de 72.04%. La première propagation de la classe C2 permet également d'obtenir un taux proche dans les zones difficiles 86.2%, mais cette fois avec une densité de 83.4%, soit un gain de +11.36%. Par ailleurs, avec la densité sur toute l'image est également plus élevée avec la propagation (+7.91%) pour un taux d'acceptation légèrement inférieur (-1.55%). Avec l'approche *wta*, et un seuillage à 0.8, le taux d'acceptation dans les zones difficiles est de 89.61% et la densité y est de 69.69%. La première propagation de la classe C3 permet également d'obtenir un taux proche dans les zones difficiles 90%, mais cette fois avec une densité de 66%, soit un gain de +12.02%. Par ailleurs, la densité sur toute l'image est également plus élevée avec la propagation (+6.21%) pour un taux d'acceptation légèrement inférieur (-0.78%). Avec l'approche *wta*, et un seuillage à 0.85, le taux d'acceptation dans les zones difficiles est de 92.59% et la densité y est de 33.12%. La

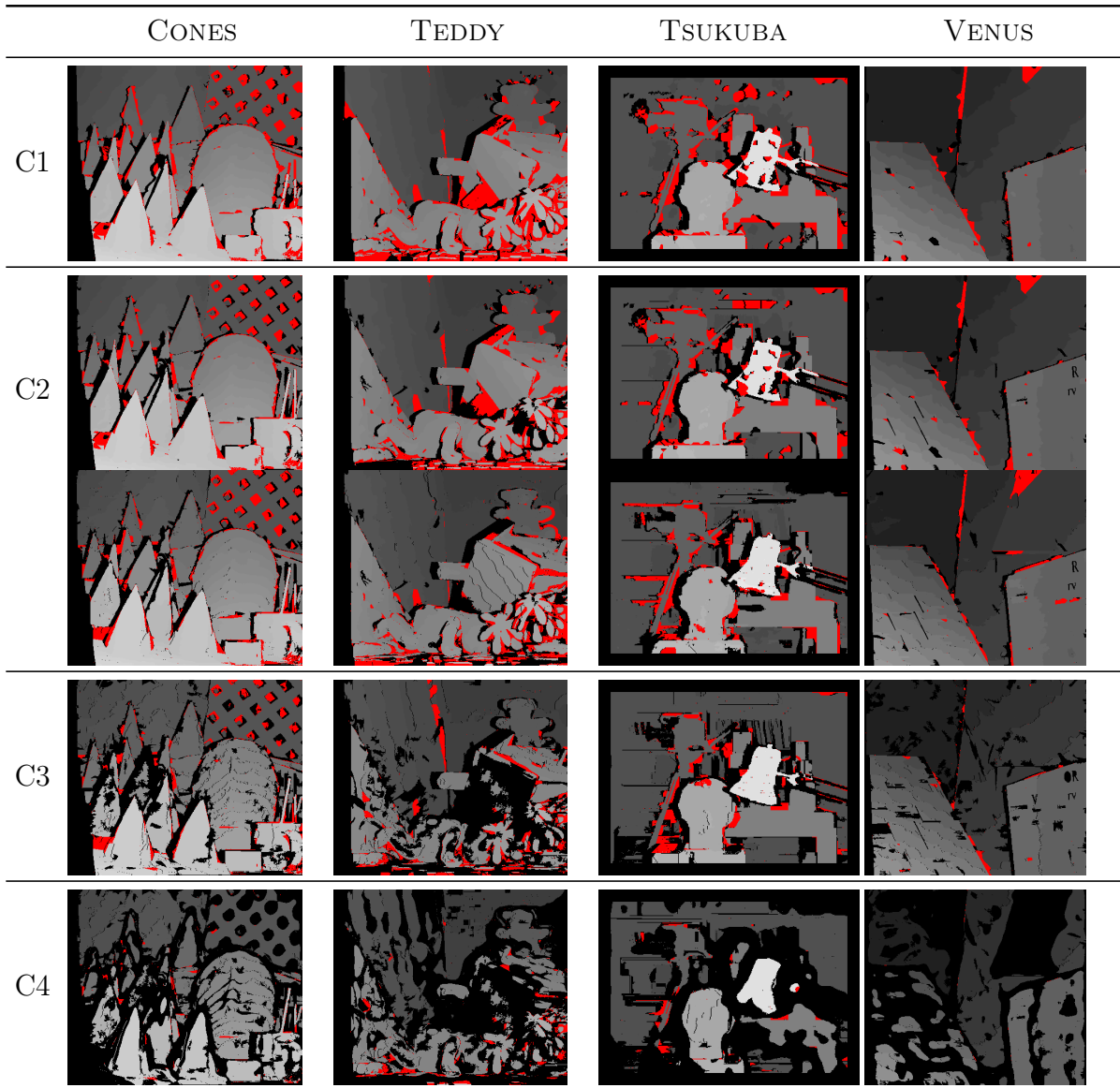


FIGURE 5.8: *Cartes des disparités propagées.* – On montre la première propagation de chaque classe ainsi que la première propagation multi-mesure (de la classe C2). On marque en rouge les disparités erronées.

treizième propagation de la classe C3 permet également d'obtenir un taux proche dans les zones difficiles 92.7%, mais cette fois avec une densité de 45.6%, soit un gain de +12.48%. Par ailleurs, la densité sur toute l'image est également plus élevée avec la propagation (+6%) pour un taux d'acceptation légèrement inférieur (-0.65%).

Ces résultats montrent donc que la mise en correspondance par propagation de germes permet d'obtenir de meilleurs résultats dans les zones difficiles que l'approche locale classique tout en donnant de bons résultats sur l'ensemble de l'image.

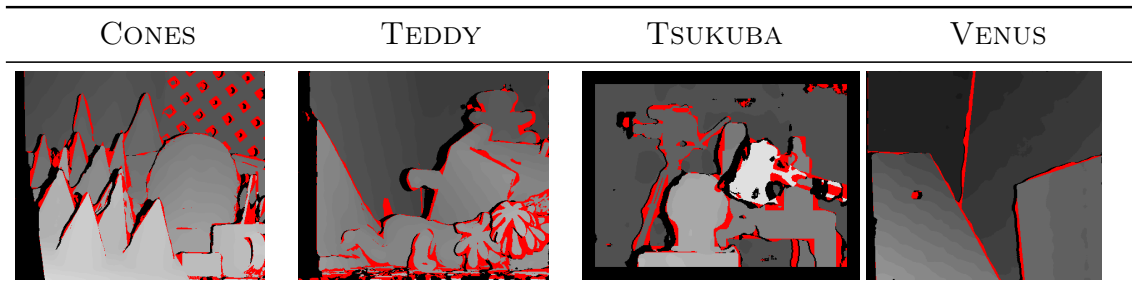


FIGURE 5.9: Cartes de disparité obtenues par mise en correspondance locale classique. – On marque en rouge les disparités erronées.

S	C_C (%)	D (%)	T_a (%)	D_{ZD} (%)	T_{aZD} (%)
\emptyset	91.96±3.03	91.07± 3.45	92.84±2.91	89±2.65	82.52±3.35
0.6	91.29±2.68	89.57±3.02	93.02±2.85	87.8±2.14	82.71±3.45
0.7	89.84±1.44	86.09±1.96	93.58±2.69	83.21±1.39	83.73±3.75
0.75	87.37±1.02	79.79±2.72	94.95±2.3	72.04±1.81	86.42±3.79
0.8	83.09±1.44	69.69±3.32	96.48±2	53.98±2.93	89.61±4.25
0.85	76.27±1.8	54.8±3.63	97.74±1.69	33.12±3.9	92.59±4.38

TABLE 5.4: Résultats de la mise en correspondance locale classique (wta) avec la mesure CENSUS.

5.4 Conclusion

Nous avons proposé une technique pour compléter et valider les germes utilisés pour la propagation qui permet d'augmenter la répartition des germes tout en conservant le taux de rappel de l'ensemble initial. Ces germes donnent de bons résultats pour la propagation, en particulier l'ensemble F (il s'agit de l'union des 9 ensembles de germes les plus complémentaires obtenus avec différents couples détecteur-corrélation paramétrés afin d'obtenir les germes donnant les taux de rappel les plus élevés) complété par les appariements d'intérêt obtenus avec la mesure $SMPD_2$ et validé. Il s'agit de l'ensemble qui a le meilleur compromis A et le taux de rappel le plus élevé. Nous avons également proposé une approche multi-mesure utilisant une mesure classique pour commencer la propagation dans les zones « faciles » d'abord puis une mesure robuste pour continuer la propagation dans les zones difficiles.

Nous avons montré que, pour trouver les meilleurs compromis avec les densités les plus fortes, l'approche simultanée avec la mesure CENSUS, sans contrainte de seuil, ni de région (ou alors avec une segmentation grossière) permet d'obtenir les meilleurs résultats. La table 5.5 résume les résultats pour les différentes méthodes locales. Les différentes classes sont déterminées en fonction des résultats obtenus. Pour obtenir un compromis avec un taux d'acceptation élevé et une densité élevée, l'approche simultanée avec CENSUS ou l'approche séquentielle avec la technique multi-mesure GC-CENSUS permettent d'obtenir les meilleurs résultats. Le seuillage sur la mesure GC doit être relativement strict et celui sur CENSUS tolérant (ou inexistant). Les erreurs sont limitées grâce à la contrainte de région qui utilise une segmentation moyenne. Nous avons également montré que les propagations séquentielles avec la mesure CENSUS seule permettent d'obtenir de bons résultats dans les zones difficiles. Enfin, pour obtenir les taux d'acceptation les plus élevés, mais avec une densité relativement faible (moins de 50%), l'approche séquentielle avec CENSUS et un seuil strict permettent d'obtenir les meilleurs résultats.

	$D \geq 80\%$	$50\% < D < 80\%$	$D \leq 50\%$
$T_a \geq 94\%$	<ul style="list-style-type: none"> ■ Propagation SE ▶ GC-C. ▶ (Strict, Tolérant) ▶ {A, B, C, D ou F}+S. ▶ S_2 	<ul style="list-style-type: none"> ■ Propagation SE ▶ GC-C. ou C. ▶ ([Strict], Strict) ▶ {A, D ou F}+S. ▶ S_1 	<ul style="list-style-type: none"> ■ Propagation SE ▶ C. ▶ (Strict) ▶ {B, D ou E}+S. ▶ S_\emptyset
$92\% < T_a < 94\%$	<ul style="list-style-type: none"> ■ Propagation SE ou SI ▶ GC-C. ou C. ▶ ([Strict], \emptyset) ▶ {C, D ou F}+S. ▶ S_2 ■ Wta ▶ (Pas de seuil ou seuil tolérant) 		
$T_a \leq 92\%$	<ul style="list-style-type: none"> ■ Propagation SI ▶ C. ▶ (\emptyset) ▶ F+S. ▶ S_\emptyset ou S_3 		

TABLE 5.5: *Bilan des méthodes locales.* - On indique, pour chaque cas, la méthode qui semble la plus appropriée et le paramétrage à effectuer (entre parenthèses on indique le seuillage à réaliser pour chaque mesure ; entre crochets, on indique les paramètres optionnels).

6 Mise en correspondance par sondage de régions

LA mise en correspondance par propagation étudiée dans les chapitres précédents donne de bons résultats quand on cherche un compromis entre le taux d'acceptation et la densité. Plus le taux d'acceptation est élevé, moins denses sont les résultats. L'objectif de ce chapitre est de proposer une méthode qui permet d'obtenir des taux d'acceptation élevés mais qui donne des résultats denses. Selon le protocole d'évaluation de Middlebury¹ qui compare les méthodes de mise en correspondance dense pour la stéréovision binoculaire, la plupart des méthodes qui sont bien classées sont fondées sur les régions. Leur principe est présenté § 2.4, page 68. Ces méthodes consistent à calculer une carte de régions de l'image en faisant l'hypothèse que chaque région de couleur homogène est la projection d'une même surface de la scène. Puis, les paramètres d'un modèle de surface sont estimés pour chaque région. Enfin, une étape finale impliquant une optimisation globale est réalisée pour raffiner les résultats. Une telle approche permet de calculer des disparités pour les zones homogènes et d'en estimer une pour les zones occultées. Néanmoins, les disparités calculées dépendent du modèle de surface choisi, le plus courant étant le modèle plan. Ce choix peut ne pas toujours constituer une bonne approximation, par exemple si un plan est choisi pour représenter une surface conique. De plus, l'étape d'optimisation est difficile car le coût utilisé est relativement complexe puisqu'il prend en compte de nombreuses contraintes pour modéliser au mieux le problème de la mise en correspondance. C'est pourquoi, nous proposons une méthode fondée sur les régions dont le principe repose sur un processus aléatoire et itératif qui calcule pour chaque pixel, à partir de plusieurs segmentations, d'une carte initiale de disparités et d'un modèle de surface, différentes approximations de sa valeur de disparité. Chaque approximation constitue un vote et la valeur qui obtient le plus de suffrages est sélectionnée comme disparité finale. Cette technique permet d'obtenir des résultats satisfaisants sans avoir recours à une étape de

1. <http://vision.middlebury.edu/stereo/>

raffinement par optimisation globale.

6.1 Algorithme proposé

L'approche proposée est résumée par la figure 6.1. Les différentes étapes sont :

- **Segmentations** – Nous faisons l'hypothèse que chaque région de couleur homogène est la projection d'une même surface et donc que les pixels d'une même région ont des disparités proches. Les méthodes fondées sur les régions utilisent généralement une sur-segmentation, c'est-à-dire de petites régions. En effet, plus les régions sont petites, plus elles peuvent être approchées par un plan sans commettre une grande erreur. Cependant, plus les régions sont petites, plus il est difficile d'obtenir des disparités initiales correctes dans les zones proches des occultations et des discontinuités de profondeur. Cela pose un problème à l'approche proposée qui s'appuie sur ces disparités. Pour le résoudre, nous utilisons différents niveaux de segmentation. Les segmentations plus fines permettent d'obtenir des valeurs de disparité approchées plus précises dans les régions où les pixels ont des disparités initiales calculées de façon acceptable. Les segmentations grossières permettent d'obtenir des votes dans les zones difficiles et occultées. En effet, les pixels situés dans ces zones ont plus de risques que les autres de ne pas avoir de disparité initiale calculée ou bien d'en avoir une erronée. Cependant, plus la segmentation est grossière, plus on augmente les chances de regrouper ces pixels avec ceux situés dans des zones qui ne présentent pas de difficulté particulière et pour lesquels la disparité initiale est acceptable. Néanmoins, on augmente également le risque de regrouper des pixels qui n'appartiennent pas à une même surface (mais ce phénomène est limité par les segmentations plus fines).
- **Mise en correspondance initiale** – Elle permet d'obtenir la carte de disparité qui est utilisée pour calculer les paramètres des différents plans pour chaque région. Cette carte doit *a priori* présenter un bon compromis entre densité et taux d'acceptation. La densité élevée permet d'avoir davantage de disparités à sonder pour obtenir un nombre plus important de votes. Le taux d'acceptation élevé permet de réduire le nombre de votes menant à des résultats erronés.
- **Sondage** – Pour chaque région de chaque niveau de segmentation, un triplet de points $\{[i_1 \ j_1 \ d_1]^T, [i_2 \ j_2 \ d_2]^T, [i_3 \ j_3 \ d_3]^T\}$ est tiré au sort dans l'espace des disparités initiales (les pixels qui n'ont pas de disparité initiale calculée ne sont pas pris en compte dans le tirage au sort). Les paramètres $[p_1 \ p_2 \ p_3]^T$ du plan sont ensuite calculés tels

que :

$$\begin{bmatrix} i_1 & j_1 & 1 \\ i_2 & j_2 & 1 \\ i_3 & j_3 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}. \quad (6.1)$$

Ce principe est illustré sur la figure 6.2 pour l'exemple d'une région représentant un cône. À chaque tirage, 3 points marqués en bleu sont tirés au sort et le plan passant par ces trois points permet de calculer une disparité approchée pour chaque pixel de la région. La carte de disparité obtenue est montrée sur la troisième colonne où les disparités erronées sont marquées en rouge. Le processus est répété plusieurs fois pour chaque région des différentes segmentations. On constate qu'à chaque tirage, une grande partie des disparités approchées est correcte. L'idée est de combiner ces différentes cartes approchées pour obtenir la carte de disparité finale. Ensuite, chaque pixel de la région reçoit un vote pour la disparité donnée par :

$$v_n = p_1 k + p_2 l + p_3 \quad (6.2)$$

où v_n est la valeur de disparité du $n^{\text{ième}}$ vote pour le pixel de coordonnées $[k \ l]^T$. Le processus est répété m fois pour obtenir m votes. Le tirage au sort peut être réalisé de façon :

- *Uniforme* – Tous les triplets ont la même probabilité d'être tirés au sort.
- *Biaisée* – Une valeur de confiance est calculée pour chaque triplet. Cette valeur détermine la probabilité d'acceptation de chaque point. Cela permet de limiter le nombre de sélections de triplets contenant un point dont la disparité initiale est erronée. La mesure de confiance, notée ct , de chaque pixel dépend de la similarité de chaque pixel et de ses voisins avec leurs correspondants estimés :

$$ct : \mathbb{N}^2 \rightarrow \mathbb{R}$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \mapsto \left(1 - \frac{\sum_{k=-r}^r \sum_{l=-r}^r \left| I^g \left(\begin{bmatrix} i+k \\ j+l \end{bmatrix} \right) - I^d \left(\begin{bmatrix} i+k \\ j+l+d_{i+k,j+l} \end{bmatrix} \right) \right|}{z} \right)^p \quad (6.3)$$

où r détermine les dimensions du voisinage considéré, z est un terme de normalisation et p est un paramètre de pondération. Plus p est élevé, plus le tirage est biaisé en faveur des pixels qui ont les valeurs de confiance les plus élevées. Néanmoins, si p est trop élevé, on risque de toujours tirer au sort les mêmes triplets. Par ailleurs, nous proposons également pour chaque région de ne conserver pour le tirage au sort, les $h\%$ de pixels les plus fiables selon la mesure de confiance (les choix de n , p et h seront discutés § 6.2.1.4).

- **Fusion** – Enfin, pour chaque pixel, la valeur de disparité qui obtient le plus de votes lui est attribuée. Pour calculer cette valeur, et pour avoir une disparité au sous-pixel, nous l’estimons en examinant le mode principal de la fonction de densité obtenue par estimation par noyau en prenant le noyau d’Epanechnikov. Il s’agit d’une technique d’estimation présentée dans [Chen 02] :

$$d_{k,l} \leftarrow \underset{v=1,\dots,d_{\max}}{\operatorname{argmax}} \frac{1}{m} \sum_{n=1}^m \begin{cases} \frac{3}{4}(1 - \|v - v_n\|^2) & \text{si } \|v - v_n\| \leq 1 \\ 0 & \text{sinon.} \end{cases} \quad (6.4)$$

La fusion est illustrée sur la dernière ligne de la figure 6.2.

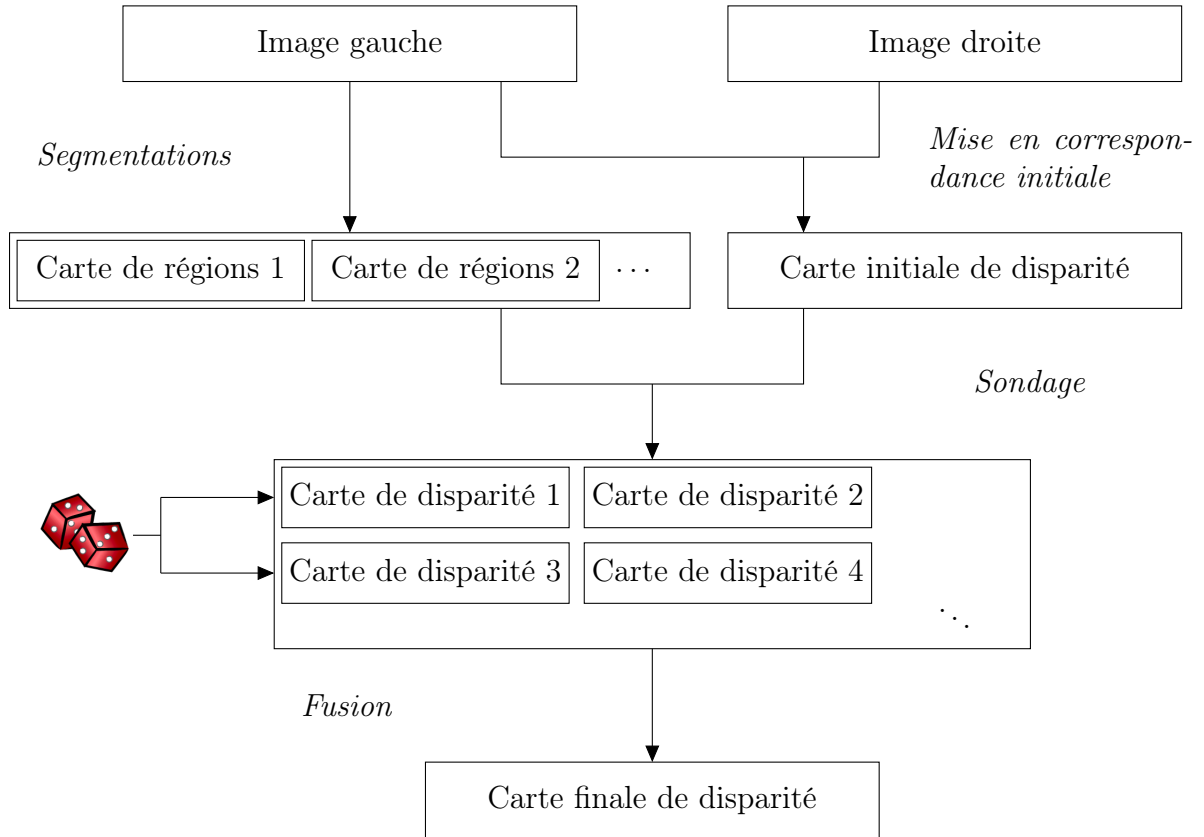


FIGURE 6.1: Étapes de la mise en correspondance par sondage de régions.

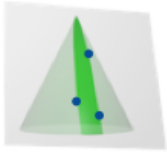
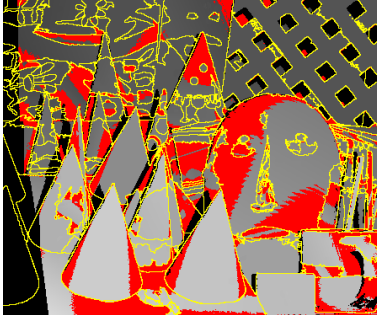
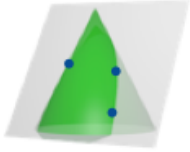
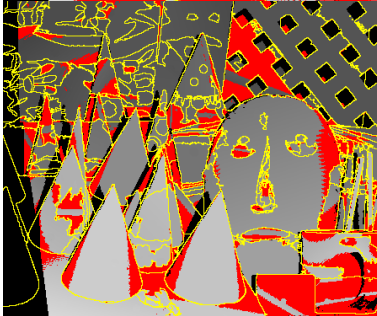
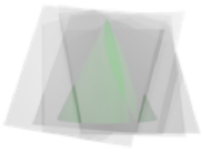
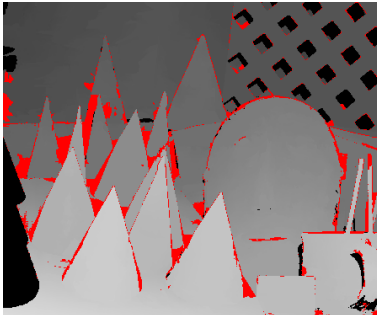
TIRAGE	EXEMPLE	DISPARITÉS APPROCHÉES
1		
2		
...
		

FIGURE 6.2: *Principe du sondage de régions.* – Sur la colonne centrale, on montre les plans induits (en gris) par les triplets (en bleu) tirés au sort pour une région représentant un cône. La dernière colonne montre les disparités approchées induites par un triplet tiré au sort par région. Les disparités erronées apparaissent en rouge.

6.2 Évaluation du sondage de régions

6.2.1 Protocole expérimental

6.2.1.1 Images testées

Les images testées sont les mêmes que dans les chapitres précédents, voir § 4.2.1.1, page 108.

6.2.1.2 Critères d'évaluation

Pour évaluer les résultats du sondage, nous utilisons les critères suivants :

- **Taux d'acceptation** T_a et **taux d'acceptation dans les zones difficiles** T_{aZD} – Il s'agit des mêmes taux que ceux utilisés § 5.2, page 136.

REMARQUE : bien que le résultat de la mise en correspondance par sondage de régions soit supposé être dense, il se peut que dans certaines régions situées dans les zones occultées, aucune disparité initiale n'ait pu être calculée, et donc que les pixels de la région soient non appariés après sondage. Nous avons donc choisi que ces taux ne soient évalués que sur les disparités calculées.

- **Pourcentage de mauvais pixels** BP (*bad pixels*) – Il s'agit du critère d'évaluation du protocole de Middlebury [Scharstein 02] :

$$BP(\mathcal{M})_\tau = \frac{\text{card} \left\{ d_{i,j} \in \mathcal{M} \mid \text{Err} \left(\begin{bmatrix} i \\ j \end{bmatrix} \right) > t \right\}}{\text{card } \tau} \quad (6.5)$$

où $\tau = \{no, all, zd\}$ est un sous-ensemble de pixels de l'image (*no* : pixels non-occultés, *all* : tous les pixels de l'image, *zd* : pixels dans les zones proches des discontinuités) et t est un seuil. Ce taux est calculé pour chacun des 3 sous-ensembles. De plus, nous calculons la moyenne de ce taux sur ces 3 ensembles et sur les 4 couples d'images, ce qui donne une appréciation sur la qualité des résultats. On note BPM ce pourcentage moyen.

6.2.1.3 Paramètre fixé

Nous fixons les différentes segmentations utilisées au nombre de 4. Nous choisissons une segmentation grossière ($h_s = 2, h_r = 12$), 2 segmentations moyennes ($h_s = 12, h_r = 6$

et $h_s = 8, h_r = 4$) et une segmentation fine ($h_s = 12, h_r = 2$). Ces segmentations ont donné de bons résultats sur les images testées lors de nos tests préliminaires. En outre, nous avons pu constater que le choix des différentes segmentations ne fait pas varier les résultats de manière significative. Toutefois, si de grosses erreurs de segmentations sont présentes sur une carte (en particulier pour la segmentation grossière), il est impératif que celles-ci ne réapparaissent pas sur les autres segmentations.

6.2.1.4 Paramètres testés

Nous faisons varier les paramètres suivants :

- **Carte initiale** – Nous testons la méthode avec la carte calculée par CENSUS selon la méthode *wt* et une contrainte de seuil (fixé à 0.7, voir table § 5.4), les premières cartes obtenues par propagation pour chaque classe (voir table § 5.3) et les cartes obtenues par propagation de la classe C2 (c'est-à-dire celles qui présentent un bon compromis entre densité et taux d'acceptation, voir § 5.3.1.4).
- **Sondage** – Nous testons le sondage avec un tirage au sort uniforme et un tirage au sort biaisé. D'après des tests préliminaires, nous avons fixé les paramètres décrits § 6.1. Nous prenons $r = 1, p = 10$ et nous testons $h \in \{35\%, 50\%, 75\%\}$. Nous faisons également varier le nombre de tirages pour chaque région $n \in \{5, 25, 75\}$. La figure 6.3 montre la carte de confiance pour le couple *Cones* où, plus le niveau de gris d'un pixel est clair, plus il a de chances d'être tiré au sort. Cette figure présente également les histogrammes des valeurs de confiance pour $p = 1$ et $p = 10$. Avec $p = 1$, la plupart des pixels ont une valeur de confiance très élevée et donc quasiment la même probabilité d'être tirés au sort. Avec $p = 10$, les valeurs de confiance sont beaucoup mieux réparties dans l'intervalle $[0; 1]$.

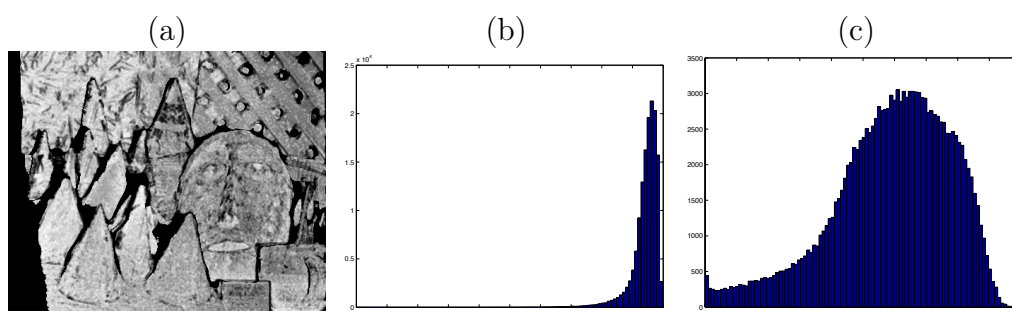


FIGURE 6.3: Carte (a) et histogrammes des valeurs de confiance pour $p = 1$ (b) et $p = 10$ (c).

n	CARTE INITIALE	p	h	T_a	T_{aZD}
75	CENSUS	10	35%	94.3±3.3	90.1±5.2
25	CENSUS	10	50%	94.2±3.4	90.0±5.5
75	CENSUS	10	50%	94.2±3.4	89.9±5.3
25	CENSUS	10	35%	94.2±3.4	89.9±5.2
75	CENSUS	10	75%	93.9±3.6	89.4±5.5
75	C1	10	50%	93.8±4.7	89.7±7.8
25	CENSUS	10	75%	93.8±3.6	89.2±5.6
25	C1	10	0.5	93.7±4.7	89.5±7.9
75	C1	10	35%	93.7±4.8	89.6±7.6
25	C1	10	0.75	93.7±4.8	89.4±7.8

TABLE 6.1: Résultats de la mise en correspondance par sondage de régions. – Tri par T_a .

6.2.2 Résultats

Résultats

La table 6.1 présente les 10 meilleurs résultats moyens obtenus sur les images testées. Plus le nombre de tirages par région est important, meilleurs sont les résultats. Cependant, on constate qu'il est possible d'obtenir de bons résultats à partir de 25 tirages. La carte initiale calculée avec CENSUS et la méthode *wta* permet d'obtenir de bons résultats. Il en est de même pour la première carte propagée de la classe C1, voir § 5.3.1.4. Il s'agit de cartes de disparités présentant un bon compromis entre la densité et le taux d'acceptation mais où la densité est un peu plus importante que les autres. On constate également que le tirage au sort biaisé permet d'obtenir les meilleurs résultats. Les meilleurs taux d'acceptation obtenus sont de l'ordre de 94% sur toute l'image et de 90% dans les zones difficiles. Il s'agit de taux élevés mais, cette fois, avec une densité maximale.

Comparaison au classement de Middlebury

La table 6.2 présente les pourcentages de mauvais appariements en fonction de t pour les meilleurs résultats que nous avons obtenus. La méthode proposée est actuellement compétitive avec celles qui sont classées sur le site de Middlebury, en particulier quand on considère une précision au sous-pixel. À titre indicatif, la méthode par sondage se positionne dans le premier quart du classement lorsque les résultats sont évalués au sous-pixel ($t \leq 0.75$) et dans la première moitié dans les autres cas.

Sur la figure 6.4, on montre les meilleurs résultats où les disparités erronées sont marquées

t	0.5	0.75	1	1.5	2
<i>BPM</i>	12.4	9.09	7.08	5.89	5.14

TABLE 6.2: *Pourcentages moyens de mauvais appariements. – Ce taux est calculé pour chacun des 3 sous-ensembles no, all, zd. De plus, nous calculons la moyenne de ce taux sur ces 3 ensembles et sur les 4 couples d’images, ce qui donne une appréciation globale sur la qualité des résultats.*

en rouge pour différentes valeurs du seuil t .

6.3 Conclusion

Dans ce chapitre, nous avons proposé une méthode permettant d’obtenir des cartes de disparité denses en s’appuyant sur différents niveaux de segmentation en régions de couleur homogène. Une carte de disparité initiale calculée avec une méthode locale est sondée pour chaque région par tirage au sort et, pour chaque pixel, la valeur de disparité qui a reçu le plus de votes est attribuée. Cette méthode donne de bons résultats, notamment avec une précision au sous-pixel. Par ailleurs, même si les tirages aléatoires ne garantissent pas que l’on obtienne les mêmes résultats quand l’algorithme est relancé plusieurs fois, avec un nombre de tirages supérieur à 75, on obtient à chaque fois des résultats très proches. Ainsi, à ce stade, nous pensons que la mise en correspondance par sondage de régions constitue une bonne stratégie de mise en correspondance dense pour la stéréovision bino-culaire. Nous avons montré qu’en moyenne, les meilleurs résultats sont obtenus à partir de la carte de disparité calculée par la mesure de corrélation CENSUS et l’approche *winner-take-all*. Cependant, la mise en correspondance par propagation peut également donner de bons résultats, avec des temps de calcul inférieurs, si celle-ci est quasi-dense. Ainsi, nous proposons les étapes suivantes :

- **Sélection de germes** – Le détecteur FAST, associé aux mesures CENSUS ou GC, est celui qui a donné les meilleurs résultats pour la sélection des germes. La meilleure stratégie consiste à détecter et mettre en correspondance les points d’intérêt de telle sorte que les germes obtenus soient très fiables, même si leur répartition n’est pas maximale. Il est alors possible d’améliorer cet ensemble de germes, en le fusionnant à d’autres ensembles issus d’autres couples détecteur-corrélation complémentaires. En outre, les étapes de complètement, par appariement d’intérêt avec la mesure SMPD₂, et de validation permettent d’améliorer les ensembles de germes en vue de la propagation.

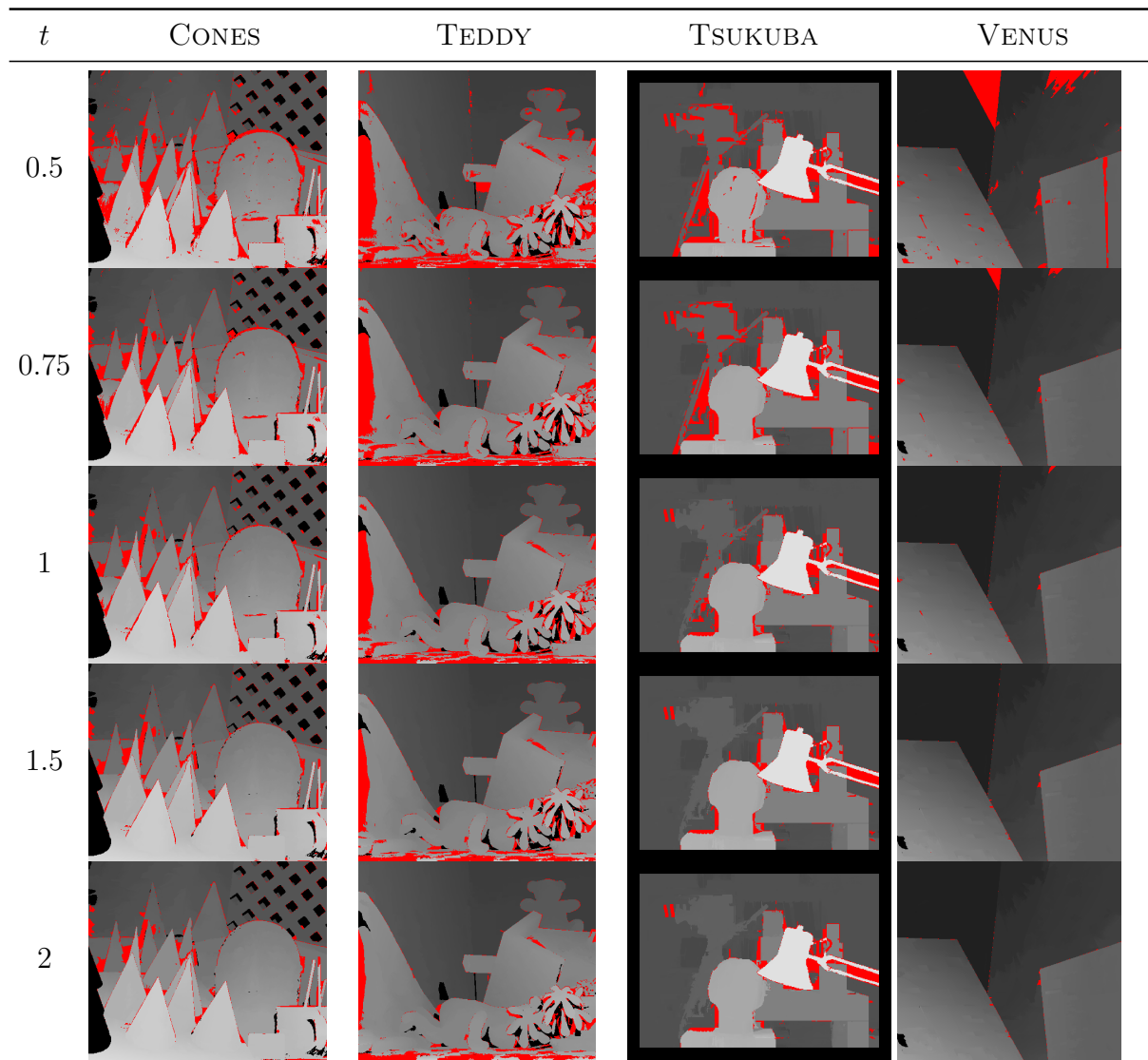


FIGURE 6.4: *Cartes de disparité obtenues par sondage des régions. – On marque en rouge les disparités erronées pour différentes valeurs de t .*

- **Mise en correspondance par propagation** – Pour pouvoir être utilisée comme carte initiale pour la méthode de sondage de régions, la carte obtenue doit présenter un bon compromis entre la densité et le pourcentage d'appariements acceptés.
- **Mise en correspondance dense par sondage de régions** – Enfin, la mise en correspondance dense peut être obtenue par sondage de régions en biaisant le tirage au sort des disparité initiales afin de favoriser les pixels qui ont les disparités initiales les plus fiables.

7 Conclusions et perspectives

7.1 Bilan des méthodes de mise en correspondance

Nous avons présenté les quatre grandes approches de mise en correspondance de pixels pour la stéréovision binoculaire. Nous proposons, dans la table 7.1, un bilan des avantages et des inconvénients de chacune des familles de méthodes.

MÉTHODE	AVANTAGES	INCONVÉNIENTS
LOCALES	Simplicité de mise en place	Sensibilité aux zones difficiles
PROPAGATION	Réduction des ambiguïtés et des calculs	Choix des germes initiaux
GLOBALES	Prise en compte de la globalité de l'information pour assurer la cohérence du résultat	Difficulté dans le paramétrage des termes de la fonction de coût
FONDÉES SUR LES RÉGIONS	Meilleurs résultats selon le protocole d'évaluation de Middlebury	Nombreux paramètres

TABLE 7.1: *Bilan des méthodes de mise en correspondance.*

7.2 Bilan des contributions

Les contributions de cette thèse portent sur les points suivants. Tout d’abord, nous avons étudié l’étape de sélection de germes pour la mise en correspondance par propagation dans le cadre de la stéréovision binoculaire. Pour cela, nous avons réalisé une analyse et une comparaison de différents détecteurs de points d’intérêt associés à différentes mesures de corrélation qui nous a conduits d’une part, à établir un état de l’art de différents détecteurs de points d’intérêt en « mettant à plat » leurs différents paramètres afin de pouvoir les comparer, et d’autre part, à définir les critères à satisfaire. Dans le cadre de la sélection de germes, il s’agit du taux de rappel et de la répartition. Plus précisément, nous avons cherché à satisfaire un compromis entre les deux. Dans ce but, nous avons déterminé les paramètres des différents couples détecteur-corrélation afin de satisfaire au mieux ce compromis. Nous avons étudié la complémentarité des différents ensembles de germes afin de les combiner pour obtenir de nouveaux ensembles satisfaisant davantage le compromis recherché. Nous avons également défini des critères à satisfaire pour la mise en correspondance par propagation, en particulier un compromis entre la densité et le taux de rappel. Nous avons proposé une approche multi-mesure qui permet de satisfaire ce compromis en tirant profit des avantages de chacune des mesures utilisées. Enfin, nous avons également proposé une méthode de mise en correspondance fondée sur les régions qui a pour but de donner un résultat dense et de bonne qualité.

De cette étude, nous donnons des indications sur les choix des méthodes de mise en correspondance qui nous semblent les plus judicieux selon des besoins spécifiques. La mise en correspondance par propagation constitue une bonne alternative à la mise en correspondance locale classique puisque les ambiguïtés et les calculs sont réduits. Cependant, elle permet d’obtenir de bons résultats à partir de germes obtenus par mise en correspondance de points d’intérêt mais au sacrifice de densité. Toutefois, nous avons réussi à obtenir une bonne densité par propagation en utilisant un ensemble initial de germes après plusieurs étapes : union de différents ensembles de germes, complètement et validation. Mais cette technique implique des calculs supplémentaires. Enfin, les meilleurs résultats sont obtenus par sondage, mais cette méthode nécessite de calculer une mise en correspondance initiale. Ces meilleurs résultats sont obtenus à partir de cartes calculées avec l’approche classique en utilisant la mesure CENSUS. Néanmoins cette mesure requiert plus de temps de calculs (de l’ordre de 6 à 10 fois plus) que les autres mesures présentées dans cette thèse. Par conséquent, en fonction des contraintes de rapidité, l’utilisateur est libre de choisir une autre mesure, mais nous pensons qu’il est important d’avoir de bons résultats dans les zones difficiles. Ainsi nous proposons comme alternative d’utiliser la mesure SMPD₂ ou bien une technique de mise en correspondance par propagation multi-mesure. Ces remarques sont résumées dans la table 7.2.

BESOIN	MÉTHODE PRÉCONISÉE
Rapidité	Propagation séquentielle de germes FAST+GC
Rapidité et densité	WTA avec mesure de corrélation GC, contrainte de symétrie souple
Densité	Propagation multi-mesure séquentielle de germes FAST+GC puis complètement avec SMPD ₂ et validation
Forte densité	Sondage de régions initié par SMPD ₂ , 3 niveaux de segmentation et environ 50-75 tirages
Meilleurs résultats	Sondage de régions initié par CENSUS, 4 niveaux de segmentation et environ de 100 tirages

TABLE 7.2: *Bilan des résultats issus de l'étude proposée.*

7.3 Perspectives

Plusieurs pistes peuvent être envisagées pour améliorer et compléter les travaux proposés dans cette thèse.

Tout d'abord, nous avons pu constater que les meilleurs germes sont obtenus en combinant des ensembles issus de plusieurs détecteurs. Nous pensons qu'un détecteur engendré par apprentissage à l'instar de FAST, pourrait être une piste intéressante à explorer.

De plus, nous avons utilisé la méthode *mean-shift* pour réaliser les différentes segmentations en régions de couleur homogène. Nous pensons qu'une étude plus approfondie sur le choix des segmentations serait pertinente. Par exemple, il peut être intéressant de combiner une segmentation basée sur la couleur avec une autre basée sur la texture.

Par ailleurs, nous avons utilisé le modèle plan pour calculer les différentes cartes de disparité approchées dans la méthode par sondage de régions, mais il peut être intéressant d'étudier d'autres types de surface, comme les surfaces *B-Spline* par exemple.

Enfin, il pourrait également être intéressant, toujours par apprentissage, d'affiner les résultats sur les différents paramétrages à effectuer avec une base de cartes théoriques plus importante et regroupées en différentes familles selon différents types d'applications visées. En effet, les applications de la vision par ordinateur sont de plus en plus nombreuses, en particulier avec l'essor récent du Web et des applications mobiles.

Bibliographie

- [Abdel-Hakim 06] A. ABDEL-HAKIM et A. FARAG. *CSIFT: A SIFT Descriptor with Color Invariant Characteristics*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, p. 1978–1983, New-York, NY, États-Unis, 2006.
- [Ayache 89] N. AYACHE. Vision stéréoscopique et perception multisensorielle. InterÉditions, 1989.
- [Babaud 86] J. BABAUD, A. P. WITKIN, M. BAUDIN et R. O. DUDA. *Uniqueness of the Gaussian Kernel for Scale-Space Filtering*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1): 26–33, 1986.
- [Banks 01] J. BANKS et M. BENNAMOUN. *Reliability Analysis of the Rank Transform for Stereo Matching*. *IEEE Transactions on Systems, Man, and Cybernetics*, 31(6): 870–880, 2001.
- [Bartoli 03] A. BARTOLI. *Reconstruction et alignement en vision 3D : points, droites, plans et caméras*. Thèse de doctorat, Institut National Polytechnique de Grenoble, septembre 2003.
- [Bay 06] H. BAY, A. ESS, T. TUYTELAARS et L. V. GOOL. *SURF: Speeded Up Robust Features*. *Computer Vision and Image Understanding*, 110(3): 346–359, 2006.
- [Beaudet 78] P. R. BEAUDET. *Rotationally invariant image operators*. Dans International Conference on Pattern Recognition, p. 579–583, Kyoto, Japon, 1978.
- [Bleyer 05] M. BLEYER et M. GELAUTZ. *A layered stereo matching algorithm using image segmentation and global visibility constraints*. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(3): 128–150, mai 2005.
- [Bleyer 10] M. BLEYER, C. ROTHER et P. KOHLI. *Surface Stereo with Soft Segmentation*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 1570–1577, San Francisco, CA, États-Unis, juin 2010.
- [Bobick 99] A. BOBICK et S. INTILLE. *Large Occlusion Stereo*. *International Journal of Computer Vision*, 33(3): 181–200, 1999.

- [Bocquillon 08] B. BOCQUILLON. *Contributions à l'autocalibrage des caméras : modélisations et solutions garanties par l'analyse d'intervalle*. Thèse de doctorat, Université de Toulouse, octobre 2008.
- [Čech 09] J. ČECH. *Accurate and Robust Stereoscopic Matching in Efficient Algorithms*. Phd thesis, Université technique de Prague, République Tchèque, février 2009.
- [Chambon 05] S. CHAMBON. *Mise en correspondance stéréoscopique d'images couleur en présence d'occultations*. Thèse de doctorat, Université Paul Sabatier, Toulouse, décembre 2005.
- [Chambon 11] S. CHAMBON et A. CROUZIL. *Similarity measures for image matching despite occlusions in stereo vision*. *Pattern Recognition*, 44: 2063–2075, 2011.
- [Chen 99] Q. CHEN et G. MEDIONI. *A volumetric Stereo Matching Method: Application to Image-Based Modeling*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, p. 1029–1034, Fort Collins, États-Unis, juin 1999.
- [Chen 02] H. CHEN et P. MEER. *Robust Computer Vision through Kernel Density Estimation*. Dans European Conference on Computer Vision, volume 1, p. 236–250, Copenhagen, Danemark, mai–juin 2002.
- [Comaniciu 97] D. COMANICIU et P. MEER. *Robust analysis of feature spaces: color image segmentation*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 750–755, San Juan, Porto Rico, juin 1997.
- [Crouzil 96] A. CROUZIL, L. MASSIP-PAIHES et S. CASTAN. *A New Correlation Criterion Based on Gradient Fields Similarity*. Dans International Conference on Pattern Recognition, volume 1, p. 632–636, Vienne, Autriche, août 1996.
- [Crouzil 97] A. CROUZIL. *Perception du relief et du mouvement par analyse d'une séquence stéréoscopique d'images*. Thèse de doctorat, Université Paul Sabatier, Toulouse, septembre 1997.
- [Devernay 94] F. DEVERNAY et O. D. FAUGERAS. *Computing differential properties of 3-D shapes from stereoscopic images without 3-D models*. Rapport de recherche RR-2304, Institut National de Recherche en Informatique et en Automatique, INRIA, 1994.
- [Fofi 04] D. FOFI, T. SLIWA et Y. VOISIN. *A comparative survey on invisible structured light*. Dans Machine Vision Applications in Industrial Inspection, volume 5303, p. 90–98, San Jose CA, USA, janvier 2004.
- [Ford 57] L. R. FORD et D. R. FULKERSON. *A Simple Algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem*. *Canadian Journal of Mathematics*, 9: 210–218, 1957.
- [Freeman 00] W. T. FREEMAN, E. PASZTOR et O. T. CARMICHAEL. *Learning Low-Level Vision*. *International Journal of Computer Vision*, 40(1): 25–47, 2000.

- [Fua 93] P. FUA. *A parallel stereo algorithm that produces dense depth maps and preserves image features*. *International Journal of Machine Vision and Applications*, 6: 35–49, 1993.
- [Gales 09] G. GALES, A. CROUZIL et S. CHAMBON. *Approches simultanées et séquentielles de la mise en correspondance par propagation*. Dans Actes du congrès francophone de Vision par Ordinateur, Objectif : Robot Autonome et Système Intelligent Sensoriel, ORASIS, support électronique, Trégastel, juin 2009.
- [Gales 10a] G. GALES, A. CROUZIL et S. CHAMBON. *Complementarity of feature point detectors*. Dans International Conference on Computer Vision Theory and Applications, support électronique, Angers, mai 2010.
- [Gales 10b] G. GALES, A. CROUZIL et S. CHAMBON. *Détection de points d'intérêt pour la mise en correspondance par propagation*. Dans actes du Congrès Reconnaissance des Formes et Intelligence Artificielle, RFIA, support électronique, Caen, janvier 2010.
- [Gales 10c] G. GALES, A. CROUZIL et S. CHAMBON. *A Region-Based Randomized Voting Scheme for Stereo Matching*. Dans International Conference on Visual Computing, volume 6454 de *Lecture Notes in Computer Science*, p. 182–191, novembre 2010.
- [Garcia 01] D. GARCIA. *Mesures de formes et de champs de déplacements tridimensionnels par stéréo-corrélation d'images*. Thèse de doctorat, Institut National Polytechnique de Toulouse, décembre 2001.
- [Geman 84] S. GEMAN et D. GEMAN. *Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6): 721–741, 1984.
- [Gevers 97] T. GEVERS et A. SMEULDERS. *Color Based Object Recognition*. Dans International Conference on Image Analysis and Processing, volume 1, p. 319–326, Florence, Italie, septembre 1997.
- [Gong 03] M. GONG et Y. H. YANG. *Fast stereo matching using reliability-based dynamic programming and consistency constraints*. Dans IEEE International Conference on Computer Vision, volume 1, p. 610–617, Nice, octobre 2003.
- [Gong 07] M. GONG, R. Y. YANG, L. WANG et M. GONG. *A Performance Study on Different Cost Aggregation Approaches Used in Real-Time Stereo Matching*. *International Journal of Computer Vision*, 75(2): 283–296, 2007.
- [Gonzalez 04] R. C. GONZALEZ, R. E. WOODS et S. L. EDDINS. *Digital image processing using Matlab*. Pearson Prentice Hall, 2004.
- [Gouet 00] V. GOUET, P. MONTESINOS, R. DERICHE et D. PELÉ. *Évaluation des détecteurs de points d'intérêt pour la couleur*. Dans actes du Congrès Reconnaissance des Formes et Intelligence Artificielle, RFIA, p. 257–266, Paris, février 2000.
- [Gruen 85] A. W. GRUEN. *Adaptive least squares correlation: a powerful image matching technique*. volume 14, p. 175–187, 1985.

- [Harris 88] C. HARRIS et M. STEPHENS. *A Combined Corner and Edge Detector*. Dans Alvey Vision Conference, p. 147–151, Manchester, Royaume-Uni, janvier 1988.
- [Hartley 03] R. HARTLEY et A. ZISSERMAN. *Multiple view geometry*. Cambridge University Press, seconde édition édition, 2003.
- [Harvent 10] J. HARVENT. *Mesures de formes par corrélation multi-images : application à l'inspection de pièces aéronautiques à l'aide d'un système multi-caméras*. Thèse de doctorat, Université de Toulouse, novembre 2010.
- [Hirschmüller 02] H. HIRSCHMÜLLER, P. INNOCENT et J. GARIBALDI. *Real-Time Correlation-Based Stereo Vision with Reduced Border Errors*. *International Journal of Computer Vision*, 47(1–3): 229–246, 2002.
- [Hirschmüller 07] H. HIRSCHMÜLLER et D. SCHARSTEIN. *Evaluation of Cost Functions for Stereo Matching*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 1–8, Minneapolis, MN, États-Unis, juin 2007.
- [Holland 75] J. HOLLAND. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence*. MIT Press, 1975.
- [Hong 04] L. HONG et G. CHEN. *Segment-based Stereo Matching Using Graph Cuts*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, p. 74–81, Washington, DC, États-Unis, juin 2004.
- [Horaud 93] R. HORAUD et O. MONGA. *Vision par ordinateur, outils fondamentaux*. Hermès, 1993.
- [Huber 81] P. J. HUBER. *Robust statistics*. J. Wiley and Sons, 1981.
- [Jurie 04] F. JURIE et C. SCHMID. *Scale-invariant shape features for recognition of object categories*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, p. 90–96, Washington, DC, États-Unis, juin 2004.
- [Kadir 04] T. KADIR, A. ZISSERMAN et M. BRADY. *An affine invariant salient region detector*. Dans European Conference on Computer Vision, volume 1, p. 228–241, Prague, République Tchèque, mai 2004.
- [Kanade 94] T. KANADE et M. OKUTOMI. *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9): 920–932, 1994.
- [Kannala 07] J. KANNALA et S. S. BRANDT. *Quasi-Dense Wide Baseline Matching Using Match Propagation*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 1–8. Minneapolis, États-Unis, juin 2007.
- [Kirkpatrick 83] S. KIRKPATRICK, C. D. GELATT et M. P. VECCHI. *Optimization by Simulated Annealing*. *Science*, 220(4598): 671–680, 1983.
- [Kirkpatrick 84] S. KIRKPATRICK. *Optimization by Simulated Annealing: Quantitative Studies*. *Journal of Statistical Physics*, 34(5–6): 975–986, 1984.

- [Kitchen 82] L. KITCHEN et A. ROSENFELD. *Gray level corner detection*. *Pattern Recognition Letters*, 1(2): 95–102, 1982.
- [Klaus 06] A. KLAUS, M. SORMANN et K. KARNER. *Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure*. Dans International Conference on Pattern Recognition, volume 3, p. 15–18, Hong Kong, août 2006.
- [Kostková 02] J. KOSTKOVÁ et R. ŠÁRA. Disparity component matching revisited. Rapport Technique CTU–CMP–2002–08, Université technique de Prague, 2002.
- [Lhuillier 02] M. LHUILLIER et L. QUAN. *Match Propagation for Image-Based Modeling and Rendering*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8): 1140–1146, 2002.
- [Li 94] S. Z. LI. *Markov Random Field Models in Computer Vision*. Dans European Conference on Computer Vision, volume 2, p. 361–370, Stockholm, Suède, mai 1994.
- [Lin 03] M. H. LIN et C. TOMASI. *Surfaces with Occlusions from layered Stereo*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8): 1073–1078, 2003.
- [Lindeberg 90] T. LINDBERG. *Scale-Space for Discret Signals*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3): 234–254, 1990.
- [Lindeberg 93] T. LINDBERG. *Detecting Salient Blob-Like Image Structures and Their Scales with a Scale-Space Primal Sketch: A Method for Focus-of-Attention*. *International Journal of Computer Vision*, 11(3): 1074–1130, 1993.
- [Lindeberg 98] T. LINDBERG. *Feature Detection with Automatic Scale Selection*. *International Journal of Computer Vision*, 30(2): 79–116, 1998.
- [Lotti 94] J. L. LOTTI et G. GIRAUDON. *Correlation algorithm with adaptive window for aerial image in stereo vision*. Dans European Symposium on Satellite Remote Sensing, p. 2315–2325, Rome, septembre 1994.
- [Lowe 99] D. G. LOWE. *Object Recognition from Local Scale-Invariant Features*. Dans IEEE International Conference on Computer Vision, volume 2, p. 1150–1157, Corfu, Grèce, septembre 1999.
- [Ma 05] Y. Z. MA et W. J. LIU. *Progressive Matching Based on Segmentation for 3D reconstruction*. Dans International Conference on Computer and Information Technology, p. 575–579, Shangai, République Populaire de Chine, septembre 2005.
- [Mikolajczyk 01] K. MIKOLAJCZYK et C. SCHMID. *Indexing based on scale invariant interest points*. Dans IEEE International Conference on Computer Vision, volume 1, p. 525–531, Vancouver, VC, États-Unis, juillet 2001.
- [Mikolajczyk 02] K. MIKOLAJCZYK. *Detection of local features invariant to affine transformations. Application to matching and recognition*. Thèse de doctorat, Institut National Polytechnique de Grenoble, juin 2002.
- [Mikolajczyk 04] K. MIKOLAJCZYK et C. SCHMID. *Scale & Affine Invariant Interest Point Detectors*. *International Journal of Computer Vision*, 60(1): 63–86, 2004.

- [Mikolajczyk 05] K. MIKOLAJCZYK et C. SCHMID. *A Performance Evaluation of Local Descriptors*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10): 1615–1630, 2005.
- [Montesinos 98] P. MONTESINOS, V. GOUET et R. DERICHE. *Differential Invariants for Color Images*. Dans International Conference on Pattern Recognition, volume 1, p. 838–840, Brisbane, Australie, août 1998.
- [Moravec 77] H. P. MORAVEC. *Toward automatic visual obstacle avoidance*. Dans International Joint Conferences on Artificial Intelligence, volume 2, p. 584–584, Cambridge, MA, États-Unis, août 1977.
- [Moravec 80] H. P. MORAVEC. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. Phd thesis, Université de Stanford, Stanford, CA, États-Unis, septembre 1980.
- [Noble 88] J. A. NOBLE. *Finding corners*. *Image and Vision Computing*, 6(2): 121–128, 1988.
- [Otto 89] G. P. OTTO et T. K. W. CHAU. “Region-growing” algorithm for matching of terrain images. *Image and Vision Computing*, 7(2): 83–94, 1989.
- [Parisot 09] P. PARISOT. *Suivi d’objets dans les séquences d’images de scènes déformables : de l’importance des points d’intérêt et du maillage 2D*. Thèse de doctorat, Université de Toulouse, janvier 2009.
- [Park 07] J. PARK, W. KIM et K. M. LEE. *Stereo Matching Using Population-Based MCMC*. Dans Asian Conference on Computer Vision, volume 2, p. 560–569, , Japon, novembre 2007.
- [Pearl 82] J. PEARL. *Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach*. Dans National Conference on Artificial Intelligence, p. 133–136, Pittsburgh, PA, États-Unis, août 1982.
- [Raoui 09] Y. RAOUI, M. DEVY et E. H. BOUYAKHF. *Scale and Rotation Invariant Local Feature Using Harris Laplace Detector in Color Textured Images*. Dans International Conference: Sciences of Electronics, Technologies of Information and Telecommunication, support électronique, mars 2009.
- [rob 94] Le Robert & Collins Senior. Harper Collins, 1994.
- [Rosten 06] E. ROSTEN et T. DRUMMOND. *Machine learning for high-speed corner detection*. Dans European Conference on Computer Vision, volume 1, p. 430–443, Graz, Autriche, mai 2006.
- [Rousseeuw 92] P. J. ROUSSEEUW et C. CROUX. *L_1 -Statistical Analysis and Related Methods*. Dans Y. DODGE, éditeur, *Explicit Scale Estimators with High Breakdown Point*, p. 77–92. Elsevier, Amsterdam, Hollande, 1992.
- [Roy 99] S. ROY. *Stereo Without Epipolar Lines: A Maximum-Flow Formulation*. *International Journal of Computer Vision*, 34(2–3): 147–161, 1999.

- [Scharstein 02] D. SCHARSTEIN et R. SZELISKI. *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*. *International Journal of Computer Vision*, 47(1): 7–42, 2002.
- [Scharstein 03] D. SCHARSTEIN et R. SZELISKI. *High-Accuracy Stereo Depth Maps Using Structured Light*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, p. 195–202, Madison, WI, États-Unis, juin 2003.
- [Scharstein 07] D. SCHARSTEIN et C. PAL. *Learning conditional random fields for stereo*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 1–8, Minneapolis, MN, États-Unis, juin 2007.
- [Schmid 00] C. SCHMID, R. MOHR et C. BAUCKHAGE. *Evaluation of interest point detectors*. *International Journal of Computer Vision*, 37(2): 151–172, 2000.
- [Shi 94] J. SHI et C. TOMASI. *Good features to track*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 593–600, Seattle, WA, États-Unis, juin 1994.
- [Smith 97] S. M. SMITH et J. M. BRADY. *SUSAN – A New Approach to Low Level Image Processing*. *International Journal of Computer Vision*, 23(1): 45–78, 1997.
- [Sun 02] J. SUN, H.-Y. SHUM et N.-N. ZHENG. *Stereo Matching Using Belief Propagation*. Dans European Conference on Computer Vision, p. 450–452, Copenhagen, Danemark, mai 2002.
- [Sun 05] J. SUN, S. B. KANG et H. Y. SHUM. *Symmetric stereo matching for occlusion handling*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, p. 399–406, San Diego, CA, États-Unis, juin 2005.
- [Sur 10] F. SUR, N. NOURY et M. O. BERGER. Determining point correspondences between two views under geometric constraint and photometric consistency. Rapport de recherche RR-7246, Institut National de Recherche en Informatique et en Automatique, INRIA, 2010.
- [Szeliski 90] R. SZELISKI. *Bayesian Modeling of Uncertainty in Low-Level Vision*. *International Journal of Computer Vision*, 5(3): 271–302, décembre 1990.
- [Szeliski 10] R. SZELISKI. *Computer vision: Algorithms and applications*. Springer, 2010.
- [Taguchi 08] Y. TAGUCHI, B. WIBURN et C. L. ZITNICK. *Stereo reconstruction with mixed pixels using adaptive over-segmentation*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 1–8, Anchorage, AK, États-Unis, juin 2008.
- [Trajkovic 98] M. TRAJKOVIC et M. HEDLEY. *Fast Corner Detection*. *Image and Vision Computing*, 16(2): 75–87, 1998.
- [Trémeau 04] A. TRÉMEAU, C. FERNANDEZ-MALOIGNE et P. BONTON. *Image numérique couleur – De l’acquisition au traitement*. Dunod, 2004.

- [Tuytelaars 08] T. TUYTELAARS et K. MIKOLAJCZYK. *Local Invariant Feature Detectors: A Survey*. *Foundations and Trends® in Computer Graphics and Vision*, 3(3): 177–280, 2008.
- [Veksler 99] O. VEKSLER. *Efficient Graph-Based Energy Minimization Methods in Computer Vision*. PhD thesis, Cornell University, Ithaca, NY, États-Unis, août 1999.
- [Vincent 01] E. VINCENT et R. LAGANIÈRE. *Matching Feature Points in Stereo Pairs: A Comparative Study of Some Matching Strategies*. *Machine Graphics & Vision*, 10(3): 237–259, 2001.
- [Wang 08] Z. F. WANG et Z. G. ZHENG. *A region based stereo matching algorithm using cooperative optimization*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 1–8, Anchorage, AK, États-Unis, juin 2008.
- [Xu 08] L. XU et J. JIA. *Stereo matching: an outlier confidence approach*. Dans European Conference on Computer Vision, volume 4, p. 775–787, octobre 2008.
- [Yang 08] Q. YANG, C. ENGELS et A. AKBARZADEH. *Near real-time stereo for weakly-textured scenes*. Dans British Machine Vision Conference, p. 80–87, Leeds, Royaume-Uni, septembre 2008.
- [Yang 09] Q. YANG, L. WANG, R. YANG, H. STEWÉNIUS et D. NISTÉR. *Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3): 492–504, 2009.
- [Yoon 05] K. J. YOON et I. S. KWEON. *Locally Adaptive Support-Weight Approach for Visual Correspondence Search*. Dans IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, p. 924–931, San Diego, CA, États-Unis, juin 2005.
- [Zabih 94] R. ZABIH et J. WOODFILL. *Non-parametric Local Transforms for Computing Visual Correspondence*. Dans European Conference on Computer Vision, p. 151–158, Stockholm, Suède, mai 1994.
- [Zhou 02] J. ZHOU et J. SHI. *A robust algorithm for feature point matching*. *Computer and Graphics*, 26(3): 429–436, 2002.
- [Zitnick 07] C. L. ZITNICK et S. B. KANG. *Stereo for Image-Based Rendering using Over-Segmentation*. *International Journal of Computer Vision*, 75(1): 49–65, 2007.

